Highly driven polymer translocation in the presence of external constraints: Simulations and theory

David Sean-Fortin

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Ph.D. degree in Physics

Ottawa-Carleton Institute of Physics
Department of Physics
Faculty of Science
University of Ottawa
Ottawa, Canada

Cette thèse est dédiée à Pascale

Summary

DNA sequencing via nanopore translocation was a pipedream two decades ago. Today, biotech companies are releasing commercial devices. Yet many challenges still hover around the simple concept of threading a long DNA molecule through a small nanoscopic pore with the aim of extracting the DNA's sequence along the process.

In this thesis I use computer simulations to create what are in essence virtual prototypes for testing design ideas for the improvement of nanopore translocation devices. These ideas are based on the general concept of modifying the average shape of the initial DNA conformations. This is done, for example, by introducing new geometrical features to the nanopore's surrounding or by the means of some external force.

The goal of these simulations is not just to test design improvements, but also to systematically deconstruct the physical mechanisms involved in the translocation process. The roles of pore friction, initial polymer conformations, monomer crowding on the *trans*-side of the membrane, Brownian fluctuations, and polymer rigidity can, with careful consideration, be essentially muted at will. Computer simulations in this sense play the role of a sandbox in which the physics can be tinkered with, in order to assess and evaluate the magnitude of certain approximations found in theoretical modelling of translocation. This enables me to construct theoretical models that contain the necessary features pertaining to the different designs tested by simulations.

The work presented here is thus constituted of both Langevin Dynamics simulations and adaptations of the Tension-Propagation theory of polymer translocation when the polymer is subject to the various test conditions.

Sommaire

Il y a vingt ans, le séquençage de l'ADN au moyen de translocation par nanopore n'était que chimère. Aujourd'hui, des sociétés de biotechnique produisent des dispositifs commerciaux. Mais de nombreux défis demeurent quant à la notion toute simple d'enfiler une longue molécule d'ADN à travers un pore nanoscopique dans le but d'en extraire sa séquence pendant le processus.

Dans la présente thèse, j'utilise des simulations informatiques pour créer, en essence, des prototypes virtuels afin de tester des idées pour l'amélioration de la translocation par nanopore. Ces idées sont basées sur la notion générale de modifier l'étendue moyenne des conformations initiales de l'ADN. On le fait, par exemple, en introduisant de nouvelles géométries dans l'environnement immédiate du nanopore ou au moyen de forces externes additionnelles.

Le but de ces simulations n'est pas seulement de tester des améliorations potentielles, mais aussi de déconstruire de manière systématique les mécanismes physiques impliqués dans le processus de translocation. Les rôles de la friction dûe au pore, des conformations initiales du polymère, de l'entassement des monomères sur le côté *trans* de la membrane, des grandeurs des fluctuations Browniennes et de la rigidité du polymère peuvent, avec une analyse soigneuse, être essentiellement contrôlés à volonté. C'est ainsi que les simulations jouent le rôle d'un 'bac à sable' dans lequel les différentes contributions à la physique peuvent être manipulées. Ceci, afin d'évaluer et de mesurer l'ampleur de certaines approximations que l'on trouve dans la modélisation théorique de la translocation. Je peux ainsi construire des modèles théoriques qui comprennent les caractéristiques requises par rapport aux différents conditions mis à l'épreuve par les simulations.

Le travail présenté est donc constitué par des simulations de types Dynamique de Langevin, ainsi que par des adaptations de la théorie de la propagation de la tension, de la translocation des polymères lorsque le polymère est assujetti à diverses conditions.

Statement of Originality

The research presented in the following thesis came out of my training as a PhD candidate. Before my involvement, Prof. Gary Slater was active in field of nanopore translocation. I inherited valuable insight for my work from him and his trainees.

I have published in peer review journals three papers (Chapters 2, 3 and 7), one more is under review (Chapter 4), and one (Chapter 5) is ready to be submitted. As first author in three of the above (Chapters 3, 4 and 7), all of the generated simulation data, the analysis, and the theoretical derivations are my original work.

In the two other papers (Chapters 2 and 5) where I am second author, I have contributed to the development of the ideas involved, the literature review, the manuscript, and in some cases (Chapter 5) I have re-generated the simulation data in its entirety as a cross-check with the data supplied by Hendrick de Haan. In these papers, I wrote less than half of the final text, approximately 20% and 40% for Chapters 2 and 5 (respectively). I have explicit permission by the first author, Hendrick de Haan, to include these as Chapters in this thesis.

For two of the above papers, I include contributions of additional simulations, analysis and theoretical modeling (Chapter 6 and 8) for completeness. In both cases these will be shortly written in article form to be submitted for peer review where I will be the first author.

The Langevin Dynamics simulations presented in this thesis were performed using the ESPResSo simulation package. The simulation scripts as well as the source-code modifications were written by me. The theoretical adaptations of Tension-Propagation theory found in some of the chapters are my adaptations of the theoretical framework first proposed by Prof. Takahiro Sakaue and co-workers.

Acknowledgements

Je tourne la page sur un très gros chapitre de ma vie qui a été profondément stimulant et rempli de défis à relever. Rien de tout ce que j'ai accompli n'aurait été possible sans la confiance, le soutien, l'aide et l'amitié d'un bon nombre de personnes dont certains que je tiens à remercier plus spécifiquement.

Merci Gary pour ton soutien indéfectible pendant toutes ses années, pour ta confiance en moi depuis le premier jour! Merci pour ton intelligence et ton encadrement flexible qui a su donner place à beaucoup d'inspiration et de créativité intellectuelle. Tu m'as assuré une liberté académique fondamentale si précieuse dans mon cheminement. Merci pour ton humanité et pour ton sens de l'humour incomparable!

Merci Vincent de m'avoir écouté, conseillé et appuyé dans ma démarche. Merci pour nos échanges d'idées de coin de table!

The research group has played a pivotal role in maintaining an intellectually stimulating environment—carefully orchestrated by Gary I suspect! I would like to thank Hendrick for his words of wisdom, for constantly grounding me back to my primary focus. Tyler for his infectious excitement for anything physics. Mykyta for his generously sharing his wealth of knowledge. Martin for being an endless source of stimulating discussions. Martine for demonstrating her courage and showing me that everything is possible. Max for making sure I constantly got distracted and making me feel bad about it—très bien! Le and Hanyang for helping with the boxes and messages, Panos for the tea, Seb for the random tasty treats.

À mes chers amis du voisinage qui on su être comme une deuxième famille pour nous, merci - gracias Gilles, Sylvain, Joséphine, Marie et Pato d'avoir contribué et donné si généreusement!

None of this would have been possible without the support of my friends in times of need. Vous avez été là même quand ce n'était pas évident. Sans vous mes amis, cette aventure n'aurait sûrement pas été la même. Votre amitié m'inspire et me touche. Merci Sean, Monica, Pierre, Émilie, les Francis, Mélanie, Marie, Annie et Geneviève...

Merci Maman pour ton amour, ta confiance, tes petites attentions, d'avoir été là pour les filles et de les avoir comblé d'amour et de précieux temps de qualité. Merci d'être venu donner un coup de main quand j'en avais besoin. Merci Papa d'avoir été un modèle, d'avoir cru en moi avec ton petit sourire en coin qui me pousse à toujours me dépasser. Merci pour les petites surprises gastronomiques, les visites impromptues et oui un gros merci pour ton aide avec la maison. Merci aux familles de Courville Nicol, Fortin et Sean pour votre soutien et compréhension. Jacques et Michelle, je suis chanceux de faire partie de votre belle famille. Je suis reconnaissant pour votre aide, votre disponibilité, votre générosité, votre amour. Merci Michelle d'avoir toujours été là pour les enfants d'une façon si enrichissante. Que dire pour toutes ses heures de révisions et de corrections... Anne-Sara et Vithia, je vous aime. Le foué est chanceux. J'ai hâte de pouvoir être là pour vous comme vous l'avez été pour moi!

Merci surtout à mes trois filles Loïc, Juliette et Soriya pour votre patience. J'ai rarement été là quand il le fallait. Ça n'a sûrement pas été facile pour vous. En espérant qu'il ne soit pas trop tard. Je vous aime fort fort!

Merci à toi Pascale qui a beaucoup trop donné:

—Cette thèse est autant la tienne—

Table of contents

1

Introduction: Faster, smaller, cheaper

It is impossible to ignore the digital revolution that has taken place in the last few decades. My doctoral work with Prof. Slater is a testament to that in the chosen approach to scientific discovery, the tools involved, and the subject of inquiry.

Historically, scientific research rests on the two pillars which are theoretical modelling and natural experimentation. Over the years a third paradigm has been steadily growing. The use of computer simulations creates virtual worlds, worlds that are governed by theoretical models unto which virtual measurements can be made. These worlds offer computational physicists an opportunity to test the consequences of certain approximations or simply test new ideas. Design prototypes no longer need to be physically built: proof-of-concepts are now made *in-silico*.

My simulations are conceived and designed primarily on a desktop or laptop computer, but the computations need to be performed on remote state-of-the-art machines. I cannot expect my virtual experiments to perform better than nature—the *real* experiments are carried out in a lab. To obtain results in a reasonable timescale I need to “cut corners” so-to-speak. Even with the coarse models that I use, computational resources remain of primary importance. It is no coincidence that the emergence of this third method of scientific inquiry closely follows the steady rise of computational power.

It is obvious how my chosen scientific approach and the needed tools are contingent upon the digital revolution. Perhaps less obvious is how this revolution is a precursor to my subject of study. The digital revolution has created immense commercial pressure on engineers and chip manufacturers to produce faster, smaller, and cheaper integrated circuitry. This is arming the bio-physics community with an array of tools aimed at the fabrication of sophisticated structures at the nanoscale. These new technologies play a pivotal role in the development of highly scalable and portable *lab-on-a-chip* devices which are not only shrinking the cost of assays, but are also bringing them out of the lab and into the hands of the clinician. Today, we can use these newly acquired tools to develop novel bio-analysis devices which—just like computers—are faster, smaller and cheaper. A big player in these new classes of technologies is nanopore DNA translocation, the topic of this thesis.

DNA is present in every single one of your cells. Every cell thus encodes the instructions needed to completely rebuild an identical version of yourself. Perhaps in a more practical sense, your cells also contain information about your ancestry, your predisposition (or lack of) to certain genetic diseases and even the potency that certain drugs will have on you. Lowering the cost of DNA analysis can one day lead to tailored medical treatments where not only the drug dosage, but also the type of drug used is chosen for a specific biological profile.

For beings as complex as ourselves, this blueprint must be exceedingly small if it is to fit inside every human cell—it is in fact 46 *single molecules*. It is both very small and quite long. Although these molecules can be stretched out and measure a combined length of a few metres, the fact that DNA is thin (a few nanometers) means it can be sufficiently compacted to fit inside the cell nucleus. Extracting information out of DNA entails finding how to work with these long molecular chains. Chain links of sugar and phosphate groups make up the structural backbone of DNA. Along this backbone are the encoding nucleotides adenine, cytosine, guanine, and thymine (`A`, `C`, `G` and `T` respectively). The information is stored and encoded in the identity of these four bases.

The famous Watson-Crick double helix, called the B-form, is the most common form of DNA. This is present at conditions found inside cells. A single DNA strand, called *single-stranded* DNA or ssDNA, can be zipped with a complementary strand to make *double-stranded* DNA, or dsDNA. The complimentary bases `A` and `T` form two hydrogen bonds wheras the bases `C` and `G` form three. In order to unambiguously read the base sequence, the native dsDNA form is usually denatured into its two ssDNA strands.

DNA sequencing entails identifying—in *sequential order*—about 3.2 billion monomeric units, where each measures less than a nanometre: 0.34 nm for dsDNA and 0.43 nm for ssDNA. The human genome was first sequenced in 2001 using a method developed by Frederick Sanger called Sanger sequencing. The precise methodology is outside the scope of this thesis, but behind the bio-chemical details awaits the fundamental physical step of being able to separate a mixture of DNA fragments by size with a resolution of a single base.

Complete DNA sequencing is almost always an overkill for most practical applications. In forensic DNA fingerprinting for example, it often suffices to ascertain whether an unknown DNA sample is the same or different than a known sample. In this situation the long dsDNA molecule is first cut into smaller fragments using *restriction enzymes* which cleave the DNA chain at sites where a specific sequence is found. For example, the *Eco*RI enzyme cuts the DNA chain at every occurrence of the sequence of `AATTC` found in the chain. Since this short sequence may occur at different positions for unrelated people, the size distribution of fragments may be different between DNA samples originating from different people. As with Sanger sequencing, this analytic method heavily relies on the ability to separate a mixture of DNA fragments by size.

Since DNA becomes negatively charged in solution, a naïve separation technique would drive the molecules by the means of an electrical field in an aqueous buffer solution. Longer fragments which hold a higher charge should migrate faster than smaller, less charged fragments. Of course, for separation to occur with this mechanism, the higher driving force on longer fragments needs to scale differently with DNA size than the opposing higher drag for longer fragments. Due to the nature of the electro-hydrodynamics in play, the drag on an electrically driven DNA molecule is actually proportional to the number of bases N, which happens to be the same as for the electrical driving force. The two terms cancel out and all the fragments migrate as the same rate, independent of N. Free flow DNA electrophoresis thus fails as a separation technique.

Enter gel electrophoresis. Discovered by accident, electrically driving the DNA fragments into a network of fibres which act like a sieving medium allows for smaller fragments to generally move through faster than longer fragments. The process is straight-forward: load the DNA mixture at one end of a gel slab, turn on the field for a given period of time, stain the DNA and take a picture of the resulting bands that have separated. One caveat is that there needs to be a sufficient amount of DNA material inside the gel slab in order to properly visualize the bands of stained fragments that separate in the gel.

In other words, a single DNA molecule will not generally be enough. As a consequence, there needs to be a DNA *amplification* step where the original DNA is copied multiple times in order to generate a sufficiently large population of identical DNA molecules.

Traditional *macroscopic* techniques that require populations of analytes certainly still have their place, but probes that promise finesse at the *single-molecule* level are becoming increasingly popular in both applied science and for fundamental research [1, 2]. Clinical point of care devices allow for immediate testing, thereby bypassing the step where samples are sent off to medical laboratories which later relay the results. Portable microfluidic devices that can test for a variety of electrolytes, specific cardiac markers, and a host of blood gases are now commercially available. An obvious advantage here is the use of a smaller sample volume: a single blood drop is all that is needed. If we push the reduction of a sample volume to its extreme, we face a new analysis paradigm: single-molecule analysis.

The ability to perform, for example, a DNA fingerprint profile with a single DNA molecule may initially seem attractive because it forgoes an entire step (DNA amplification) in the analysis pipeline. This is more than a mere practical advantage. The ability to perform analysis at a single-molecule level can reveal the fundamentally rich nature of macromolecules. Outlier data which are statistically absent in a large ensemble can hold important biological information. An example is the use of single-molecule *mechanochemistry* that reveal kinetic pathways of protein unfolding in ways which are not possible using an ensemble of molecules [1].

The idea that chemically identical molecules can contain a high degree of heterogeneities is called *molecular individualism*. Pierre Gilles deGennes first coined this term when considering experimental data of elongating single DNA molecules under extensional flow [3]. There are no obvious secondary or tertiary changes to take the blame here; the effect is rather a result of the vast conformational phase space accessible to macromolecules.

We do not yet have the ability to construct the molecular machinery that our bodies use to read and manipulate DNA, but we can nonetheless turn to nature for inspiration. For example, the Enterobacteria phage T2 virus infects a host cell by directly injecting its DNA through the cell's membrane by piercing it with a hollow tube-like structure. The virus' DNA can then enter the cell through this small pore. In essence, this thesis is focused on the biomimicry of this process. Recent technological advancements allow for the drilling of a nano-sized pore across a thin solid-state (usually silicon) membrane.

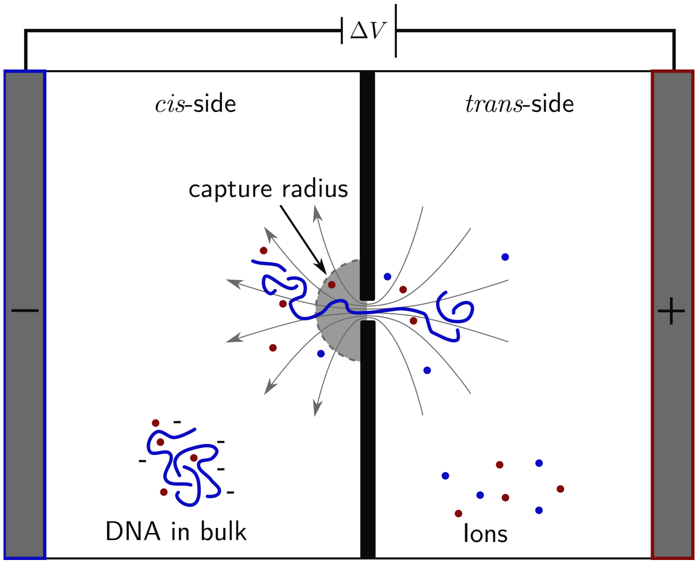


Figure 1: Schematic of a typical translocation setup. Bulk DNA diffuses in the *cis*-side region before being captured by the electric field lines that focus near the nanopore. The resulting process brings the DNA towards the *trans*-side through the nanopore. The DNA's presence in the nanopore can cause a partial blockage of ions. Note that the negatively charged DNA is driven in the direction *opposite* the direction of the field lines.

Since DNA becomes charged in solution, it can be driven from one side of the membrane to the other by the application of an external electric field.

Nanopore translocation technologies can be cheaper and faster than traditional DNA analysis techniques. Furthermore, these technologies can be integrated into high-throughput micro and nanofluidic systems. Due to the single molecule working principle, these analysis methods can be sensitive to sample heterogeneities and can track rare outlier features which can hold valuable information.

The Coulter counter can perhaps be considered as a technological predecessor of nanopore translocation. The essential concepts are the same: i) two chambers filled with a conducting solution are divided by an electrically isolating barrier which contains a small hole; ii) an electric potential difference is applied between the two chambers such that the field lines converge near the hole. The current modulation that occurs when an object obstructs the hole opening is used as the main readout. The working principle is that different sized objects modulate the current readout differently. This is primarily used as an automated way of counting red and white blood cells, replacing the time-consuming staining and manual counting steps.

George Church and David Deamer independently [4] proposed that the simple working principles of the aforementioned Coulter counter could be applied to DNA sequencing. Figure 1 depicts a schematic of two chambers divided by a thin membrane with a pore. The solution contains small mobile ions (eg., salt) which migrate towards oppositely charged electrodes by crossing through the pore. The migration of these charged species through the pore can be read as a baseline *ionic current*.

DNA, which is negatively charged in solution, is loaded in the left chamber in Fig. 1, where it randomly diffuses. We refer to this as the *cis*-side. When a DNA molecule approaches the nanopore where the density of field lines increase, the magnitude of the electric force acting on the DNA becomes sufficiently high such that the DNA motion becomes more *driven* than *diffusive*. This defines the *capture radius*. Once inside this zone, the DNA is siphoned inside the nanopore and transits towards the *trans*-side, on the right. The presence of the DNA molecule inside the pore causes a dip in the ionic current for the duration of the *translocation event*.

The original sequencing idea is that given a sufficiently small (nano-sized) pore, as the DNA transitions through the pore in a single file, the individual DNA bases could modulate the current trace such that the sequence could be interpreted from it. The nanopore would not only need to have a small diameter to enforce the single file translocation, but would also have to have a small axial dimension such that the current trace is predominately affected by a single base inside the pore.

The elegance of having the DNA driven through the pore by the same electrical potential that would offer the current readout is however deceitful. It is much more practical to have an independent control over the readout and the driving aspects of an experiment. The double edged role of the driving field here means that one cannot tune capture probability and translocation speed without affecting the quality of the signal. Moreover, even an atomically thin graphene membrane [4] may not be able to offer the needed resolution.

The ionic current of anions and cations, just like the DNA, is driven by the converging field lines in the vicinity of the nanopore. This zone can be seen as some kind of 'read-head' where DNA material affects ionic conductivity the most. The size of this zone is determined by the pore geometry such as length and diameter: the smaller these geometrical parameters are, the smaller, or more "focused" the read head. Since the smallest diameter has to be sufficiently large such that the bases can fit through, the required hole diameter ≈ 3 nm sets an upper limit. Yet even an infinitely thin membrane would yield blockaded signals modulated by more than one nucleotide [4], thus making sequencing very challenging.

This has motivated the use of small electrodes embedded inside the nanopore. The current modulation *across* these electrodes can potentially offer localized nucleotide-specific signals while the DNA translocates in a single-file manner. Embedded electrodes offer a way to de-couple the driving and reading aspects of translocation. Unfortunately, there remain two aspects of the driving field which aren't so easily de-coupled: capture and translocation rate—an experimentalist is usually left with a compromise between the two.

A DNA molecule ending up on the *trans*-side by translocation is first captured on the *cis*-side after having diffused into the zone described by a *capture radius*, see Fig. 1. The size of this zone depends upon the applied potential difference and the nanopore geometry. The aim of single-file translocation puts an upper limit on the nanopore diameter, whereas the membrane is chosen to be as thin as possible. This leaves the applied potential difference as one of the few remaining parameters to control the capture rate. With nanopore diameters of only a few nanometers, a potential difference of the order of $\sim$ 200mV is needed to create a capture radius large enough to acquire diffusing DNA at a reasonable rate. Under such potential differences DNA translocates at an average rate in the order of 1 nucleotide per μs [4]. These rates imply a needed detection bandwidth currently unattainable [4] for nucleotide-specific current modulation.

1.1 Thesis overview

This thesis is focused on the use of computer simulations as a means to explore various translocation-specific prototype scenarios. I will focus on exploiting DNA confinement under various geometries as a means to control the kinetics of translocation.

My computer simulations employ a coarse-grained model which is best interpreted as being some kind of generic polymer and nanopore. Supplemental to the simulations, a theory for driven translocation is adapted to the various scenarios investigated. Since the simulations and the theory share some of the same generic features, results from both approaches are compared to shed light on the impact of their respective associated approximations. Features which are difficult to treat analytically can often be included with minor computational costs in the simulation model, helping bridge the gap between theory and experiments.

The first article (Chapter 2) focuses on finding simulation parameters which would yield simulation dynamics consistent with what is found in the lab by matching the relevant Péclet number.

The second article (Chapter 3) looks at the effect of pre-confining the polymer in a long cylindrical tube before the translocation process takes place. As a follow-up paper (Chapter 4), we generalize the pre-confinement into a finite-sized tube and investigate the effect of changing the tube's aspect-ratio. After presenting my work during a CAP meeting in Ottawa, I've had the opportunity to submit an extended abstract to the journal *Physics in Canada*. This can be found in Appendix A, where I present the contents of the tube paper for a wider audience.

Motivated by reducing the width of the distribution of translocation times, we investigate (Chapter 5) the relative effects of changing two independent sources of noise—namely, Brownian and conformational. The latter is controlled via the use of an applied pre-stretching force. As a complementary addition to this rather empirical manuscript, I include a manuscript that focuses on a theoretical model (Chapter 6) of the same pre-stretching scenario.

Finally, we test how translocation can be slowed down by forcing the DNA into a cross-linked gel region on the *trans*-side (Chapter 7). This is followed by a short theoretical addendum (Chapter 8). A proof-of-concept of this idea was experimentally realized with the help of collaborators from the research group led by Prof. Vincent Tabard-Cossa. The paper, published in *Electrophoresis*, can be found in Appendix C for reference.

Over the course of my PhD work on polymer translocation, I have had the opportunity to contribute to additional research closely related to polymer physics and biological physics. Although these collaborations go beyond the main theme of driven polymer translocation, I include them in Appendices for completeness. These reflect my other research interests and a significant portion of my training and contributions as PhD student.

During my work with Prof. Slater I have had the opportunity to supervise a summer student. Together, Erica Wang and I developed an idea that Prof Slater and I had while I was working on my Masters degree. By the time the summer contract with Erica had ended, we had submitted a manuscript to the journal *Electrophoresis* which was eventually accepted and published. This opportunity at *Highly Qualified Personnel* training marks the beginning of my PhD, and although unrelated to driven translocation, this remains for me an important accomplishment, which represents one of many duties as a PhD student in the Slater lab. I have thus chosen to include it here as Appendix B.

I have presented the work done by Erica and I at a CAP meeting in Montreal where I also met Prof. Sabrina Leslie and became acquainted with some of her work. Prof. Leslie invited Hendrick de Haan (then a postdoctoral fellow in the Slater group) and I to visit her lab which showcases the “CLiC” apparatus: a means to continually confine DNA molecules in a quasi-parallel plate geometry across many orders of magnitude for the plate separation distance. This led to a collaborative effort of theory, simulations and experimentation on DNA confinement between two parallel walls which I have included in Appendix D. The manuscript is currently under review by the journal *Macromolecules.*

Prof. Leslie was invited to give a talk at the University of Ottawa, and this spawned a second collaborative project with Tyler Shendruk (then a PhD student in the Slater group). This project aimed at wrapping DNA around a micro-sized rod as a DNA spool. We filed an invention disclosure and produced a manuscript which I have included here in Appendix E. The manuscript is currently under review by the journal *Physical Review X*.

Having been acquainted with the plugin structure of the software ImageJ via analysis work for the Leslie lab, I wrote a plugin for PhD Candidate Sebastian Hadjiantoniou to help him analyze some of his experimental data. This started a collaboration which quickly grew to include a coarse-grained cellular aggregation model. With Maxime Ignacio (a postdoctoral researcher in the Slater lab), simulations were developed to test how the anisotropy of cellular aggregates depend upon the degree of confinement in a channel geometry. With the lab led by Prof. Andrew Pelling, we published the paper found in Appendix F in the *Journal of the Royal Society Interface*.

1.2 The physics of polymers

Like a chain constructed of repeating links, a polymer (from *polys* meaning many and *meros* meaning part) is a molecule constructed of many repeating units. In a linear chain where every link—save the first and last—is connected to two other links, the repeat units or *monomers* have connecting bonds between themselves to form a polymer. Although in general polymers do not necessarily have to to be linked in a linear fashion [1], the class of polymers considered in this thesis will be limited to linear polymers, as shown in Fig. 2

As most physicists can attest to, predicting the motion of these linked N-bodies can quickly become challenging. To make matters worse, I wish to consider the physical properties of a long molecular chain immersed in a thermal bath, where every monomer

---

[1] One can have stars, comb, bottlebrush, or ring polymers for instance.

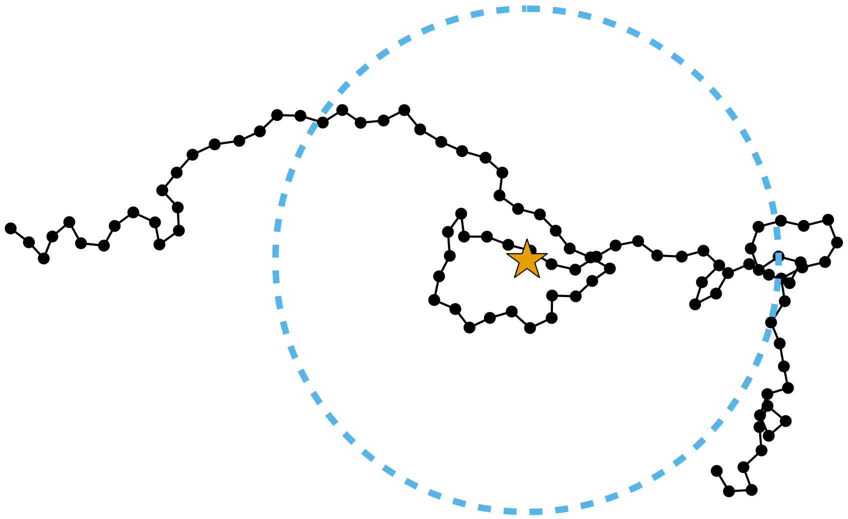


Figure 2: A two dimensional ideal linear polymer modelled by a random walk of 99 steps of fixed bond length. The star indicates the position of the centre of mass. The circle having a radius equal to the instantaneous radius of gyration, R_g, is outlined by a dashed line.

is subject to random collisions from the surrounding solvent. By admitting defeat in solving the N-body problem, one can turn to thermodynamic averaging. Analogous to the ideal gas law $PV = Nk_\mathrm{B}T$, certain properties of a single polymer chain can be surprisingly well characterized when the number of repeating units N—also called the *degree of polymerization*—is sufficiently high. In this Chapter, I will go over a few hallmark results from the application of statistical mechanics to single polymers.

1.2.1 Polymer size

In order to characterize the N-body system, we first need to define a few elementary properties such as position and size. The position of any of the monomers can be used to define the polymer's position. Whether this be the position of the first monomer, $\mathbf{r}_1$, or the last, $\mathbf{r}_N$, often makes no difference. It is however much more useful to define the polymer's position as the average position of the monomers. This metric is not only more general (useful for polymers outside the linear configuration), but intuitively seems more reasonable—this polymer metric makes use of the information contained in all of its constituents (monomers), not just one. This definition lends itself nicely to the general concept of polymer dynamics[2] which I will go over later in this chapter. As such, the polymer position is defined as

$$\mathbf{r}_{\mathrm{cm}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i. \tag{1}$$

Although this is often called the *center of mass*, the average above is unweighted. Since I am mainly interested in *homopolymers* which have identical subunits, I have omitted the masses from Eq. 1.

How does one determine the "size" of a chain? The width and length are two obvious candidates, but can sometimes be misleading. Take for example the length of chromosome 1 that, if stretched out, is ≈ 8.5 cm. This *contour length*, L_{c}, does not quite indicate what we would want to use as a "size", especially considering that every human cell somehow contains in its nucleus exact copies of this DNA molecule (and 45 other chromosomes). Just like an ideal gas explores the volume of the vessel which contains it, a polymer in thermal equilibrium with a heat bath constantly moves and deforms—but unlike an ideal gas, the monomers are *connected.* The bonds are not allowed to cross therefore creating constraints in how the polymer can deform. The monomers are also limited to positions which do not violate the step-size constraints imposed by having bonds. As such, the monomers cannot expand and occupy the volume of an arbitrary large vessel. The polymer can only occupy an effective finite volume. The key assumption which needs to be made when introducing the concept of an ideal gas to students is that the gas particles are modelled as point-like particles which do not interact with one another. This is similar to the *ideal polymer* model where the monomers are modelled as point-like particles which do not interact with other monomers—except of course through their connectivity. As such, one can construct an ideal polymer of N monomers connected by $N-1$ bonds by simply generating a random walk of $N-1$ steps of unit length b. I will call the set of monomer positions $(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$ the polymer *conformation.* In this thesis, and in the field of polymer physics in general, long polymers ($N \gg 1$) typically are the main interest. Thus for conciseness (and aesthetics), I will not be making a distinction between the number of bonds ($N-1$) and the number of monomers (N), since $N \simeq N-1$ for sufficiently large N. I will thus be referring to both simply as N.

Figure 2 shows a polymer conformation of $N = 100$ monomers in two dimensions. By modelling a polymer conformation as the path outlined by a random walk, the squared distance between the two ends of the polymer, $R_{\mathrm{e}}^2 = (\mathbf{r}_1 - \mathbf{r}_N)^2$, is thus analogous to the

---

[2] Consider the difference between single monomer diffusion and the centre of mass diffusion for the polymer as a whole.

squared distance traveled by a random walker after N steps of length b. The square-root of this metric, the end-to-end distance, is perhaps the most straight-forward definition of polymer "size" for a linear polymer. Since the value of R_e changes as the polymer takes on different conformations, it is often more useful to average R_e over all possible conformations. Assuming that all conformations are equally probable, one can calculate the average distance between the two ends of a polymer by following the formalism used to calculate the root-mean-squared distance traveled by a random walker

$$\sqrt{\langle R_\mathrm{e}^2 \rangle} = bN^{1/2}. \tag{2}$$

This relation can be obtained in a "hand waving" manner by considering the polymer not as a series of monomers ($\mathbf{r}_1 ... \mathbf{r}_N$) but rather as a a series of bonds ($\mathbf{s}_1 ... \mathbf{s}_{N-1}$) of length b. In this representation the end-to-end vector can be expressed as the sum of the tangent vectors

$$\mathbf{R}_\mathrm{e} = \sum_{i=1}^{N} \mathbf{s}_i, \tag{3}$$

and the average of its square

$$\langle \mathbf{R}_\mathrm{e} \cdot \mathbf{R}_\mathrm{e} \rangle = \langle \sum_{i=1}^{N} \mathbf{s}_i \cdot \sum_{j=1}^{N} \mathbf{s}_j \rangle = \langle \sum_{i=j}^{N} s_i^2 \rangle + \langle \sum_{i \neq j}^{N} \mathbf{s}_i \cdot \mathbf{s}_j \rangle. \tag{4}$$

The last term vanishes since the average of the uncorrelated bonds is zero, and what remains is the sum of N identical terms $\langle s_i^2 \rangle = b^2$. Taking the square-root then yields Eq. 2.

For reasons similar to why I chose Eq. 1 over the position of a single monomer, there is an alternative size metric which is often preferred over R_e, that is, the radius of gyration [5]

$$R_\mathrm{g}^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_\mathrm{cm})^2. \tag{5}$$

An outline of a circle having radius R_g is shown for the conformation of Fig. 2. This metric makes use of all the other monomers (not solely the two ends), in determining the effective polymer radius. Additionally, the radius of gyration can be calculated for all polymer configurations (such as a *ring polymer* which does not have ends or a *star-polymer* which has many). As with the end-to-end distance R_e, the average value of R_g over all conformations is usually sought after. For ideal linear polymers in 3D free space, the ratio between these two metrics is constant $\langle R_\mathrm{e}^2 \rangle / \langle R_\mathrm{g}^2 \rangle$=6. However this ratio is not generally conserved when the polymer is found under geometrical confinement or is subject to the influence of external forces, as will be investigated in this thesis. Since the radius of gyration is such a useful measure for a polymer's size, its bulk value in free solution is often used as a reference. We will thus define $R_{\mathrm{g}0} = \sqrt{\langle R_{\mathrm{g}}^2 \rangle}$ as the average radius of gyration of a polymer in free solution.

Let us we go back to $\mathbf{R}_{\mathrm{e}}$ as the final position of a random walk after N-steps. If one considers a random walk of unit step length b in one dimension, the resulting probability density of being at position $\mathbf{R}_{\mathrm{e}}$ after N-steps is given by the binomial distribution. The binomial distribution for the end-to-end distance of the random walk of an ideal polymer will approach a Gaussian distribution by virtue of the *Central Limit Theorem.* This gives the following probability density for a *Gaussian polymer* [5]

$$P(\mathbf{R}_{\mathrm{e}}) = \frac{1}{\sqrt{2\pi N b^2}} e^{-\frac{R_{\mathrm{e}}^2}{2Nb^2}}, \tag{6}$$

where $\mathbf{R}_{\mathrm{e}}$ is the end-to-end vector. Since the steps in different dimensions are independent, this can be generalized for any dimensionality.

The mean value of the $\mathbf{R}_{\mathrm{e}}$ vector is obviously zero due to symmetry i.e., Eq. 6 is indeed a centred Gaussian with zero-mean. One may again turn to the root-mean-square as a physical indicator of the spatial extent. The probability density can be used to calculate directly the mean $\langle R_{\mathrm{e}}^2 \rangle$. However, there is a simple shortcut. One may simply recognize that Eq. 2 is also obtained by simple inspection of Eq. 6, since it is written in the canonical form (the numerator of the exponent's argument is twice the variance).

Having such a description of an ideal polymer is incredibly powerful. The fact that two Gaussian polymers attached by their ends is also a Gaussian polymer [5] opens the door to analytical solutions of diverse situations. But Eq. 6 offers a bit more than that. It offers a conceptually simple interpretation of polymers which allows for intuitive thought.

Consider how, in statistical mechanics, the probability of a state has an associated weight given by the Boltzmann factor $e^{-U/k_{\mathrm{B}}T}$, where U is the energy of the microstate, k_{B} is Boltzmann's constant, and T is the temperature. Inspection of Eq. 6 can thus be interpreted as having some kind of associated energy $U(R_e) = \frac{1}{2}kR_{\mathrm{e}}^2$, with

$$k = \frac{k_{\mathrm{B}}T}{Nb^2} \tag{7}$$

applied between the two polymer ends. This is effectively a *harmonic spring* which pulls the two ends together. This conceptual *spring* is entropic in nature, and will guide theoretical constructions for polymers under various conditions.

1.2.2 Real polymers

The first *refinement* to the ideal gas law that students learn is usually the *van der Waals* equation of state. In this "more physical" representation of a gas the particles are no longer point-like but rather bead-like since they occupy a volume, and the gas beads can attract themselves via a potential. I will not be discussing the effect of monomer attraction in detail, as it mainly relates to solvation properties which are irrelevant for DNA. I am mainly interested in a polymer model where the monomers occupy a small volume which cannot overlap with the volumes occupied by other monomers.

In opposition to an ideal polymer, *real polymer* models have excluded volume interactions between the monomers. The result of having this can be seen as a "swelling" effect. Consider the set of all conformations of an ideal polymer and the fact that we have $R_{\text{g0}} \propto bN^{1/2}$. If we were to draw every monomer as spherical beads with diameter b, some of the conformations will have partially overlapping monomers (by definition adjacent monomers are separated by the bond distance b and do not overlap). If all of the conformations of an ideal polymer which contain overlapping beads are eliminated from the set of (previously) "legal" conformations, the average size of the remaining conformations is going to be slightly higher (this is because the conformations containing an overlap are, on average, systematically smaller in size). It turns out that this swelling effect is not fully accounted for by the prefactor, but is rather manifested via a different *size exponent* ν, such that

$$R_{\text{g0}} \propto bN^{\nu}. \tag{8}$$

A useful approximation in d-dimensions $\nu = 3/(d{+}2)$ is often used as a rule of thumb (it is in fact exact in one and two dimensions), but I prefer the more precise value of $\nu = 0.588$ [5] when performing numerical estimates in 3D. Notable examples of the $\nu = 3/(d+2)$ rule are for $d = 1$ and $d = 4$. Consider a random walker on a one dimension lattice (can only take left and right steps of length b): the polymer self-avoidance constraint is equivalent to forbidding the walker from walking on a lattice point previously visited. If, for example, a first step is taken towards the right, the only available choice for the second step is also towards the right (a left step would bring the walker to a previously visited site). In this trivial example, the size (end to end distance for simplicity) grows linearly as $R_{\text{e}} = bN$ (in other words, $\nu = 1$) since the self-avoidance criteria forces the—not so random—walker to only step towards the right. The one-dimensional exponent of $\nu = 1$ is thus *exact*. However, with increasing dimensions, the odds of coming back to a visited site decreases. This is particularly apparent with the four-dimensional exponent or $\nu = 1/2$, which is the same for an ideal polymer. The size exponent saturates at $\nu = 1/2$ (it cannot see itself less than not at all) because the vast amount of space available in four or more dimensions dwarves the effect of re-visiting a site.

The exponent ν is often called Flory's exponent in honour of chemist Paul Flory. Flory's mean field derivation of ν simply solves for the equilibrium between two opposing effects: the aforementioned polymer swelling *vs* the entropic contraction. In a way, Flory's derivation is not self-consistent since it will be using an ideal polymer model to obtain both of these forces.

I suggested earlier that an ideal polymer effectively acts like a Hookian spring with a force constant $k = dk_\mathrm{B}T/Nb^2$ (Eq. 7 in d-dimensions). Note that this is for an *ideal polymer*—applying this at face value for a real polymer *overestimates* the contraction. Nevertheless the free energy due to expansion can be expressed as $U_\mathrm{spring} \propto R_\mathrm{e}^2 k_\mathrm{B}T/Nb^2$. With monomers each occupying a d-dimensional *volume* of b^d, one can estimate the free energy U_EV associated with the excluded volume interactions. In thermal equilibrium, the energy due to monomer overlaps is of order $k_\mathrm{B}T$ per interacting monomer. Thus for the whole polymer with N monomers [6],

$$U_\mathrm{EV} \propto N \times k_\mathrm{B}T \times (\text{Interaction Probability})\,. \tag{9}$$

The interaction probability is estimated with the mean-field monomer volume fraction of Nb^d/R_e^d. This supposes that every monomer is surrounded by a constant-density cloud of monomers inside the polymer coil having a size characterized by R_e. The free energy can now be written as

$$U_\mathrm{EV} \propto Nk_\mathrm{B}T\left(\frac{Nb^d}{R_\mathrm{e}^d}\right). \tag{10}$$

This interaction energy cannot include two connected monomers as they are positioned by definition at a constant distance b (think of the ideal polymer where this is also true). Moreover, this estimate can be seen as a worst-case scenario, since monomers positioned far from the polymer centre cannot sample a uniform monomer "cloud". Nevertheless, one can minimize the sum of the two opposing contribution $U_\mathrm{EV} + U_\mathrm{spring}$ with respect to the polymer size R_e to find

$$\begin{aligned} \frac{d}{dR_\mathrm{e}}\left(U_\mathrm{EV} + U_\mathrm{spring}\right) &= 0 \\ \frac{6R_\mathrm{e}}{Nb^2} &= \frac{dN^2b^d}{R_\mathrm{e}^d} \\ R_\mathrm{e} &\propto bN^{3/(d+2)}. \end{aligned} \tag{11}$$

By fortuitous cancelation of overestimates, this derivation provides the useful Flory size exponent approximation $\nu = 3/(d+2)$.

1.2.3 Polymer stretching

Suppose an ideal polymer is initialized as a straight line and we are interested in following how its conformations evolve in time. The value of R_e starts at $R_\mathrm{e} = L_\mathrm{c} = Nb$ (straight line) and collapses towards lower values as it randomly "wiggles" and explores conformational space. Simply due to entropy, the end-to-end distance decreases towards $R_\mathrm{e} = bN^{1/2}$ given by Eq. 2. In the "single polymer system" studied thus far there is also a limit of $R_\mathrm{e} = Nb$, but for large N this limit is hardly sampled by the system whose root-mean-squared value is closer to $bN^{1/2}$. The odds are against us ever witnessing the polymer going back to its initial conformation where $R_\mathrm{e} = Nb$.

This is similar to the picture of an ideal gas initialized in a small volume in a corner of a given box which will rapidly expand until it occupies the total volume of the box. Just like the ideal gas particles are not likely to end up at a later time in the smaller corner where they were initialized, the monomers are not likely to end up forming a straight line. Although entropy favours the exploration of space, the expansion of the gas has its limits. The reason the gas particles do not go further than the limits of the box is due to the collision on the walls which reverses the direction as they bounce off. As we know, this is called *pressure*, and the intimate relationship between the two conjugate variables volume and pressure is described with the ideal gas law. By some external means one can, using a piston for example, compress the gas back to a state similar to how it was initialized.

Similarly, by the use of some external force, one can pull the ends of the polymer towards a state resembling the initial straight line with $R_\mathrm{e} = Nb$. Just like the gas will resist the compression by the means of an increasing force exerted on the piston, the polymer will resist stretching via a *tension*. This mechanical spring behaviour is the same that was alluded to in the discussion about the *Gaussian polymer* model.

Conceptualizing the connected N-body system as a two-body *dimer* connected by a harmonic spring can help elucidate the force-extension curves for a linear polymer. At low deformation, or low applied force, the system is harmonic. However, unlike what is suggested by Eq. 6, the two ends cannot be separated indefinitely. If the bonds do not stretch, there is a hard limit at $R_\mathrm{e} = L_\mathrm{c} = Nb$. One should expect the mean extension to asymptotically approach this value at high forces. For an ideal polymer, the force-

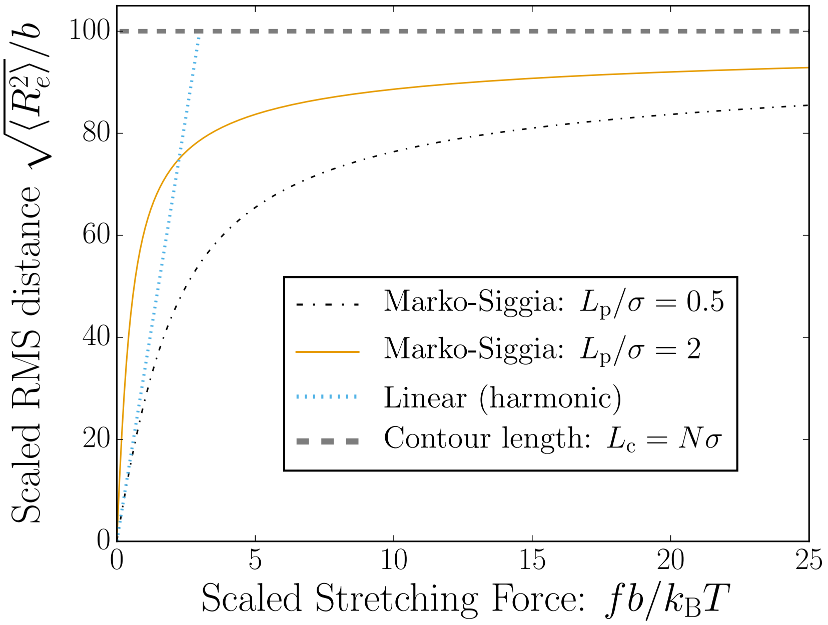


Figure 3: The root-mean-squared extension R_e *vs* the scaled force fb/k_BT as predicted by different models. The persistence length L_p, defined in the next section, describes the polymer rigidity

extension relation can be obtained analytically and is called the Langevin function [7]. However, Marko and Siggia [8] proposed an interpolation between the two regimes for the case of a semiflexible polymer. The latter model, which is better suited for DNA, shall be used in this thesis. Figure 3 shows the two limiting behaviours with the Marko-Siggia relation

$$\frac{fL_p}{k_BT} = \frac{R_e}{L_c} + \frac{1}{4(1 - R_e/L_c)^2} - \frac{1}{4}, \tag{12}$$

where R_e is the root-mean-squared extension, and L_p is a parameter describing the polymer rigidity (see next section). At low forces, the linear harmonic spring is adequate. The Marko-Siggia function breaks off the linear harmonic regime and asymptotically approaches the limit of $L_c = Nb$ for high extensions. The high extension limit is derived from the wormlike chain model that uses a continuous (smooth) representation of a polymer with a bending cost for high curvature regions. I include in Fig. 3 a force-extension curve (labeled $L_p = 2\sigma$) for a polymer that is not fully flexible, but is rather *semiflexible*—the subject of the next subsection. The cases for $L_p = 0.5\sigma$ can be considered flexible since L_p is comparable to the monomer size.

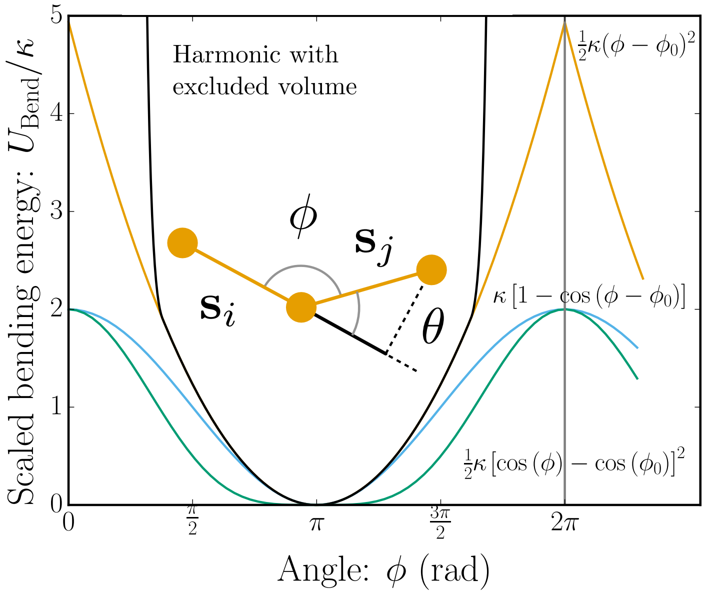


Figure 4: Various widely used bending potentials. I will be using the harmonic potential. The harmonic potential curve shown in black includes excluded volume interactions from next-nearest neighbours. In all cases $\phi_0 = \pi$. The angles are defined using a cartoon trimer. The projection of $\mathbf{s}_i \cdot \mathbf{s}_j$ is shown as a bold black line.

1.2.4 Polymer rigidity

So far, I have considered models for polymers that are flexible and can freely fold over by any amount—with the exception of some effects due to excluded-volume. In this subsection, I consider the effect of an energetic cost associated with the angle formed between two consecutive bonds. The goal is to introduce rigidity into the polymer model.

In computer simulations, the average angle formed between two consecutive bonds can be controlled by the use of an angular potential. A straight-forward choice is the harmonic potential given by

$$U_{\mathrm{bend}} = \frac{\kappa}{2}\left(\phi - \phi_0\right)^2 , \tag{13}$$

where κ is a bending constant, ϕ is the angle formed between two neighbouring bonds and ϕ_0 is the equilibrium angle (as shown in the inset of Fig. 4). Figure 4 shows the shape of U_{bend} along with other possible choices. In this thesis I focus on an equilibrium angle $\phi_0 = \pi$ such that a very high value of κ results in a rod-like polymer.

In the preceding subsection I made use of Eq. 4 and the fact that the bond vectors were uncorrelated to find $\langle R_{\mathrm{e}}^2\rangle$. The addition of a stiffness potential introduces finite correlations between the bond vectors. The correlation between two bond vectors of length b is shown in Fig. 4. One can interpret the projection $(1/b)\mathbf{s}_i \cdot \mathbf{s}_j = b\cos(\theta)$ as what is left of the original bond vector $\mathbf{s}_i$. From this the following question can be asked: "how many more bonds are needed until the memory of the first orientation is lost?" The trimer cartoon in Fig. 4 shows that the original orientation is reduced by $\cos(\theta_{ij})$. An additional projection from the next bond vector $\mathbf{s}_k$ will reduce $\cos(\theta_{ij})$ by another similar amount $\cos(\theta_{jk})$.

The memory of the original bond vector $\mathbf{s}_i$ will decay after sufficient additional projections. We can take the ensemble average and assume that a given orientation will be reduced, on average, by $\langle\cos(\theta)\rangle$ per step:

$$\begin{aligned} \frac{\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle}{b^2} &= \langle\cos(\theta)\rangle^{i-j} \\ &= e^{\ln\left[\langle\cos(\theta)\rangle^{i-j}\right]} \\ &= e^{|i-j|\ln[\langle\cos(\theta)\rangle]} \\ &= e^{-|i-j|b/L_\mathrm{p}} \end{aligned} \tag{14}$$

where the characteristic decay length

$$L_\mathrm{p} = \frac{-b}{\ln(\langle\cos(\theta)\rangle)} \tag{15}$$

is called the *persistence length* [6].

In my work I will mostly use the harmonic bending potential given in Eq. 13. In the work presented in this thesis the bond length b is almost identical to the monomer diameter such that the beads are in contact, similar to a pearl necklace. The force discontinuities arising from the kink at $\phi = 0$ are avoided since I mainly focus on real polymers with excluded-volume interactions; these interactions are felt between next-nearest neighbours when $\phi \lesssim \pi/3$ (60°) and ensure that the harmonic potential is not pathological. The data plotted in Fig. 5a shows that an exponential fit to the decay in bond vector correlations yields a characteristic length L_p/b close to the stiffness constant $\kappa/k_\mathrm{B}T$. One can associate a Boltzmann weight $e^{-U_\mathrm{Bend}/k_\mathrm{B}T}$ to all possible trimer conformations (see the inset of Fig. 4 for an example). In two dimensions, the average of $\cos(\theta)$

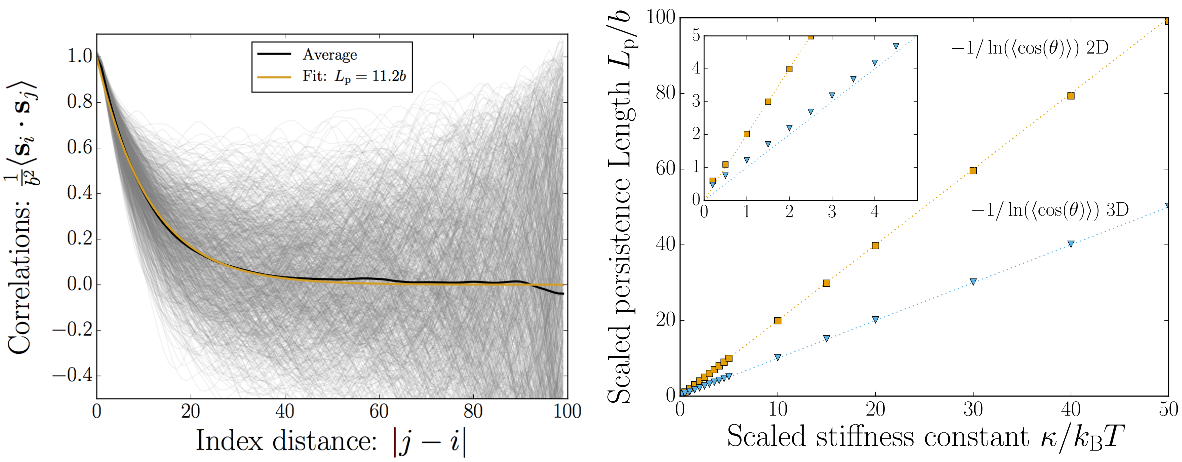


Figure 5: a) The bond vector correlations as a function of index distance $|i-j|$ along the backbone. The persistence length is calculated via a fit to simulation data of a $N = 100$ real semiflexible polymer. The grey lines correspond to 1000 independent conformations having a harmonic bending potential with $\kappa = 10k_\mathrm{B}T$. b) The scaled persistence length and a function of the scaled stiffness constant. Using the harmonic potential, the average $\langle\cos(\theta)\rangle$ is calculated both in 2D and 3D from which I obtain the persistence length via Eq 15 (shown as symbols).The dotted lines show the approximation $L_\mathrm{p}/b = \kappa/k_\mathrm{B}T$.

is found by evaluating

$$\langle\cos(\theta)\rangle = \frac{\int\limits_{\theta=0}^{\pi} \cos(\theta) e^{-\kappa\theta^2/k_\mathrm{B}T} d\theta}{\int\limits_{\theta=0}^{\pi} e^{-\kappa\theta^2/k_\mathrm{B}T} d\theta}.$$

In three dimensions, both integrals are evaluated over the solid angle $\sin(\theta)d\theta d\alpha$, where $\alpha \in (0, 2\pi)$. In Fig. 5b, the mean $\langle\cos(\theta)\rangle$ is numerically evaluated in 2D and 3D to obtain the persistence length L_p (using Eq. 15). We include in Fig. 5b the approximation $L_\mathrm{p}/bk_\mathrm{B}T \approx 2\kappa/(d-1)$, where d is the dimensionality, as a practical rule-of thumb [9]. The resulting curves are remarkably accurate down to low values of κ, shown as the inset.

We must keep in mind that the evaluation above does not take into account effects from excluded volume interactions (which, for example, change the integral bounds). For most practical values of κ, however, the energetic cost of increasing the angle θ dwarves the excluded-volume effect of next-nearest neighbours.

For $\phi_0 = \pi$ the angle θ is centred at zero, $\langle\theta\rangle = 0$. Deviations from this equilibrium

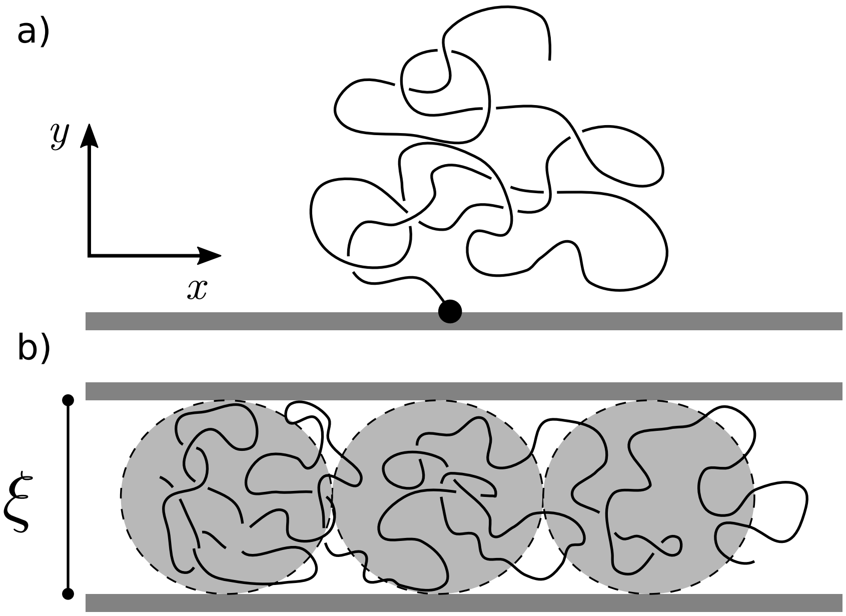


Figure 6: Two situations of a confined polymer. a) A polymer tethered to a wall. b) A polymer confined inside a tube of diameter ξ together with a representation of a blob construction.

angle are due to thermal fluctuations. For at high $\kappa \gg k_\mathrm{B}T$, the small angle approximation $\langle \cos(\theta) \rangle \approx 1 - \langle \theta^2 \rangle /2$, and $\ln(1 - \langle \theta^2 \rangle/2) \approx -\langle \theta^2 \rangle /2$; thus [6],

$$\frac{L_\mathrm{p}}{b} \approx \frac{2}{\langle \theta^2 \rangle}. \tag{16}$$

1.2.5 Polymers in confinement

When polymers are not in free-space conditions however, the bulk properties discussed above can no longer be used at face value: nevertheless, it is helpful to consider them as a reference point. The radius of gyration in bulk, $R_{\mathrm{g}0}$, for example is often used to determine the length scale when deformations arising from geometrical confinement become significant. Figure 6 shows two cases where geometrical constraints affect polymer conformations.

A tethered polymer, shown in Fig. 6a, has its first monomer fixed to a flat wall. An obvious consequence is that the mean end to end distance in the y-direction is no longer zero due to symmetry breaking. The effect of the wall can be seen as contributing to some kind of effective repulsion in the $+y$ direction applied on the monomers. The *half-space* partition function of a Gaussian polymer can be evaluated, but for more exotic scenarios where excluded volume interactions, polymer rigidity or an applied force are also involved, computer simulations are a perfectly suitable mean of providing numerical data.

The schematic in Fig. 6b shows a polymer confined inside a tube of diameter ξ. The deformation depicted is typical for $\xi < R_{\mathrm{g0}}$. Blob theory is a powerful tool for characterizing the polymer conformations under such a deformation. Consider a random walk initialized inside the tube. One may ask how many steps does it take until it collides with the tube wall? Since the end-to-end distance after g-steps scales as bg^{ν} (Eq. 8), one can guess that collisions occur when $bg^{\nu} \approx \xi$. This gives us the scaling relationship $g \sim (\xi/b)^{1/\nu}$. Below this scale the walk obeys bulk statistics and above this length scale some kind of deformation occurs. This defines the *blob* size ξ. Figure 6b shows how the deformed polymer can be represented as a series of self-avoiding blobs having size ξ. In the schematic, three blobs of size ξ are linearly arranged, which yields a polymer end-to-end distance of $R_{\mathrm{e}} \cong 3\xi$. In general, the number of blobs is simply given as N/g, so that

$$R_{\mathrm{e}} \simeq \frac{N}{g}\xi \simeq N\left(\frac{b}{\xi}\right)^{1/\nu}\xi,$$

which gives R_{e} as a function of the known quantities b, N, and ξ.

1.3 The polymer DNA

Two opposing paradigms: DNA as a model polymer *versus* a polymer model for DNA.

On the one hand, theoretical predictions made by polymer physicists were historically difficult to verify with synthetic polymers due to the difficulty in producing mono-disperse laboratory samples. For this reason, the use of the biological polymer DNA became quite appealing since there are sophisticated biological processes which ensure that DNA's are copied with little tolerance to errors. For example, λ-DNA, which contains *exactly* 48502 basepairs, can be inexpensively mass-produced. This, along with other DNA's have played an important role not only in the biological study of DNA but to also in the field of polymer physics. My experience with people who follow this paradigm is that they are more often inclined to the fundamental understanding of chain-like molecules. For them, DNA is a convenient experimental tool to test what are usually generic theories.

On the other hand, the obvious relevance of understanding DNA cannot be ignored and sophisticated DNA models are constructed by polymer physicists and chemists to help further the field. Various elaborate simulation models for DNA have been constructed in order to investigate DNA-specific phenomena. These are often supported by atomistic simulations of DNA, which share little with the long chain, long timescale limits of generic theories. My experience with scientists that belong to this group is that they are more inclined towards working for a specific application. They see polymer theories as a tool to be used on DNA as a way to develop practical applications.

This thesis will be sampling from both point of views. I will be considering the former paradigm—the polymer physicist's approach—in how the model is constructed. I will not be explicitly considering chemical interactions, but will rather focus on fundamental physical phenomena. This is the same as how the ideal gas law does not care about whether the gas is Helium or Argon. Where such details for an ideal gas are found in coefficients like the specific heat c_v, the details regarding the polymer model will be found in coefficients like persistence length L_p. On the other hand the prototypes I will be considering are focused around biotechnology. Some of the proposed systems in this thesis offer novel approaches to DNA sizing or improvements to nanopore based DNA sequencing and analysis.

In short, the work presented in this thesis follows a philosophy akin to *the development of the van der Waals gas law to build a better heat engine*.

1.3.1 Coarse-grained DNA

Partly due to stacking interactions between the bases, and partly due to its relatively high charge, dsDNA is quite rigid. Its persistence length of $L_\mathrm{p} \approx 50$ nm is more than two orders of magnitude higher than the distance between bases $b \approx 0.34$ nm.

Consider two DNA segments, either from the same molecule (spatially close but far along the DNA contour) or from two different molecules. They will be influenced by a repulsive Coulombic interaction. However, positively charged salt ions in solution are attracted to the negatively charged DNA segments and can partially screen out the repulsion between these DNA segments. Although the steric width of dsDNA is close to $w = 2$ nm, a more reasonable estimate would include effects from this Coulombic repulsion and/or a hydration layer which increases the effective width. Under high salt conditions it is estimated that $w \approx 5$ nm [10].

When the double helix is unzipped into two complementary strands, the resulting ssDNA pieces have quite different physical properties. The persistence length is drastically reduced to $L_\mathrm{p} \approx 2$ nm [10]. Interestingly, the distance between bases is *increased* to $b \approx 0.43$ nm which is close to its effective width $w \approx 0.6$ nm [10].

In the majority of cases, a bead-spring simulation model is used in this thesis. This

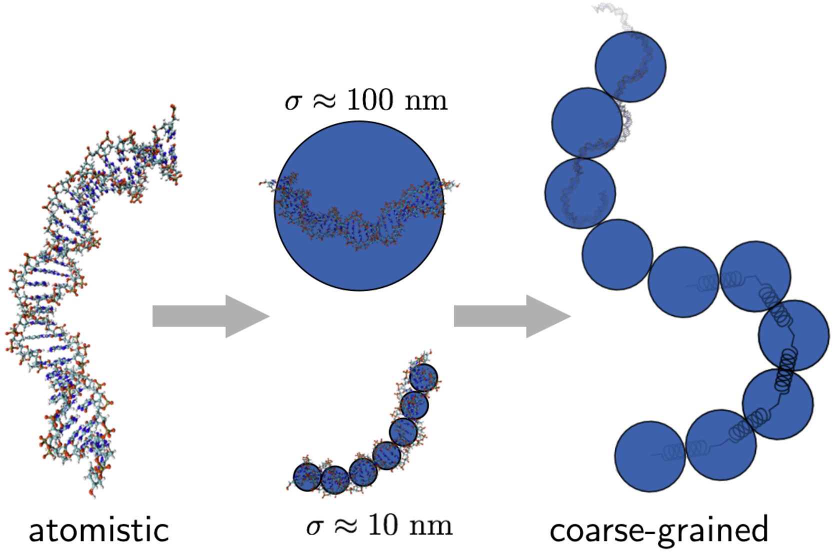


Figure 7: Coarse graining dsDNA (not shown to scale). Computationally expensive atomitic details are coarse-grained out by the use of coarse simulation beads

is described in more details in subsection 1.5 below.

From this generic polymer model, one can attempt to find a DNA equivalent *after the fact*. The length scale of a simulation bead where the polymer is fully flexible has to correspond to *at least* a flexible length of DNA. This means that a simulation bead diameter of $\sigma = 2L_{\mathrm{p}} \approx 100$ nm for dsDNA and $\sigma = 2L_{\mathrm{p}} \approx 4$ nm for ssDNA. This map will propagate to all remaining length scales which are going to be measured in units of σ, including the contour length, the polymer width, the nanopore diameter, *etc.*

Obviously, the consequence of a coarse-grained model is that some length scale ratios will not be correct. Figure 7 shows how a coarse $\sigma \approx 100$ nm representation does not capture the correct width of dsDNA. On the other hand a polymer modelled to capture the correct width $\sigma = w \approx 5$ nm will need to have an additional bending potential to match the correct persistence length as well as more beads to yield a given DNA length. Similar arguments can be made for length scales pertinent for ssDNA or the nanopore geometry.

As far as length scales are concerned, I will be focusing on a generic polymer with a generic nanopore. The model will follow closely existing coarse-grained simulation and theoretical models in the literature.

The goal is to build a simple theory which can be later generalized to include different effects. For example, in Chapters 7 and 8, I briefly explore how progressively ramping up the persistence length affects translocation. Advancing in a general progressive manner can permit one to eventually "plug-in" relevant parameters for a specific polymer (dsDNA or ssDNA).

Length scales are not the only relevant properties which plays a role in modelling polymer translocation. For the translocation process considered in this thesis, energy and time are two additional pertinent physical quantities which need to be addressed. To investigate times scales, one must consider polymer dynamics, the subject of the next section.

1.4 Polymer dynamics

Consider a spherical particle suspended in a heat bath. Random collisions between the object and the solvent will impose random forces. This can cause a directed motion or slow down an existing movement. Over the course of many random collisions, the object will follow *Brownian* motion. Such random movement of an object is what leads to *diffusion*.

The diffusion coefficient of an object following Brownian motion can be defined by how its mean-squared displacement $\langle \Delta r^2 \rangle$ grows in time

$$\langle \Delta r^2 \rangle = 2dDt, \tag{17}$$

where the prefactor depends on the dimensions d and D is the *diffusion coefficient*.

The solvent plays the role of a heat bath. We will implicitly consider it as a source of both viscous damping and random forces. A friction coefficient ζ can be seen as a damping parameter. For example, in low Reynolds number conditions of interest here, the laminar flow of solvent moving around an object having a velocity $\mathbf{v}$ will be impeding its motion via a force $\mathbf{F}_{fric} = -\zeta \mathbf{v}$. A high value of ζ will yield a higher damping force, whereas $\zeta = 0$ reflects completely invisible solvent and the object will move in ballistic, vacuum-like conditions.

On the other hand, the object's ability to sense the solvent will allow it to sample random fluctuations from the heat bath. The latter has an energy scale $\sim k_{\mathrm{B}}T$. This rather different viewpoint suggests that ζ is not only a term to incorporate viscous damping, but also one which help bring about random fluctuations.

This is an example of *fluctuation-dissipation theorem*. The way in which the solvent both impedes motion (friction) *and* causes motion (diffusion) is captured via the relationship between the coefficient ζ and the diffusion coefficient D. This is beautifully described by three simple terms in the *Einstein-Smoluchowski* relation

$$D = \frac{k_{\mathrm{B}}T}{\zeta}. \tag{18}$$

When a polymer is considered as a single object having a size R_{g0}, one can define the *diffusion time* τ_{diff} as the time needed to move—on average—a distance similar to its own size:

$$\tau_{\mathrm{diff}} \approx \frac{R_{\mathrm{g0}}^2}{D} \approx \frac{R_{\mathrm{g0}}^2 \zeta}{k_{\mathrm{B}}T}. \tag{19}$$

This quantity can be measured both in the lab and in simulations, allowing for a map between physical timescales and its associated derived units. Chapter 2 focuses on finding how the simulation model details affect the translocation time in units of τ_{diff}.

In *Rouse*-dynamics, the friction coefficient of an object having N-connected beads is simply the sum of the frictional contributions from each monomer. This yields the polymer friction $\zeta_p = N\zeta_m$, where ζ_m is the friction coefficient for a monomer. The Rouse diffusion coefficient is thus inversely proportional to N [5].

Consider two objects undergoing sedimentation: a solid sphere with radius R_{sp} and a polymer with radius $R_{\mathrm{g0}} = R_{\mathrm{sp}}$. To permit the passage of these sphere-like objects, the solvent needs to move out from the front, go around, and go back behind the sedimenting “spheres”. We can often assume Stokes drag on a sphere which has the form $\zeta = 1/(6\pi\eta R)$ where η is the solvent viscosity and R is the sphere radius [5].

In the long polymer limit ($N \gg 1$), the energy dissipation required to displace the solvent around a spherical object of size $R_{\mathrm{g0}} \approx bN^{\nu}$ is less than the sum of moving around N individual monomers of size b as predicted using the Rouse picture. The Rouse description where $\zeta_p = N\zeta_m$ suggests that the polymer having a radius similar to that of the solid sphere would move considerably slower—yet a significant portion of its volume is hollow! This absurd result highlights that the Rouse description largely overestimates the polymer friction coefficient. In fact, the rate at which the polymer sediments should at the very least have a size scaling similar to that of the solid sphere. This simple argument suggests that Rouse dynamics neglects a fundamental aspect: the long-range hydrodynamic interactions.

In *Zimm* dynamics where long range hydrodynamic interactions are considered, the preceding argument, that the Rouse friction coefficient is an overestimate, is confirmed. As it turns out, the polymer friction $\zeta_p \sim 1/R_{\mathrm{g0}}$ scaling is indeed predicted and observed [5].

The translocation dynamics considered in this thesis will not be considering long-range hydrodynamics interactions. This is justified by the presence of the membrane, which provides hydrodynamic screening. Computer simulations that use full hydrodynamics find the effect to be small [11]. Due to their negligible impact and their relatively high computational cost, long-range hydrodynamics will be neglected. Like many other authors in the field, we will thus be considering Rouse-like dynamics both in the simulation model and in the theoretical description of translocation dynamics.

1.5 Langevin Dynamics simulations

The simulation beads are integrated in time according to the Langevin equation of motion

$$m\ddot{\mathbf{r}} = \mathbf{F}^{\mathrm{C}} - \zeta\dot{\mathbf{r}} + \mathbf{F}^{\mathrm{B}}, \tag{20}$$

where $\mathbf{r}$ is the position vector of a simulation bead having a mass m, $\mathbf{F}^{\mathrm{C}}$ is the sum of all conservative forces, ζ is the monomer friction coefficient, and $\mathbf{F}^{\mathrm{B}}$ is a random Brownian force. The solvent is implicitly present via the two last terms which are connected by the fluctuation-dissipation theorem. An object moving through a fluid experiences viscous drag $-\zeta\dot{\mathbf{r}}$ which *dissipates* energy into heat. This has an associated *fluctuation* contribution of random collisions which ultimately gives rise to Brownian motion. To account for this, we use an uncorrelated random force vector $\mathbf{F}^{\mathrm{B}}$ having zero-mean and a variance of $\langle F_{\mathrm{B}}^2\rangle$ to be determined below.

Consider the mean squared displacement of a single bead subject to the Langevin equation. Over time scales such that the motion is over-damped (longer than the ballistic timescales) and if there are no additional forces $\mathbf{F}^{\mathrm{C}} = 0$, the above can be discretized (a necessary step in computer simulations) to find the mean-squared displacement

$$\langle \Delta r^2\rangle = \langle F_{\mathrm{B}}^2\rangle \Delta t^2/\zeta^2 = 6Dt, \tag{21}$$

which corresponds to the Einstein-Smoluchowski result if we choose $\langle F_{\mathrm{B}}^2\rangle = \frac{2\zeta k_{\mathrm{B}}T}{\Delta t}$ in each dimension.

1.5.1 Kremer-Grest polymer

In order to perform simulations using the general polymer picture of N beads linearly connected by springs, specific mathematical forms for the excluded-volume U_{EV} and spring potential U_{spring} need to be chosen. In the spirit of the pedagogical *box on an* *inclined plane* level of physics, the chemical details of DNA are not of interest here. A generic polymer made of beads connected by springs will be used in the simulations.

In order to consider these point monomers as *beads* that occupy space, the repulsive Weeks-Chandler-Andersen (WCA) potential is used [12, 13] to model excluded volume

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r < r_{\mathrm{c}} \\ 0 & \text{for } r \geq r_{\mathrm{c}} \end{cases}, \tag{22}$$

where r is the distance between the centre of two beads. The effective bead diameter σ and the well depth ϵ define the simulation units for length and energy. This is essentially a Lennard-Jones potential where the cutoff distance $r_{\mathrm{c}} = 2^{1/6}\sigma \approx 1.1224\sigma$ is at the well minimum such that this potential is purely repulsive.

Although this potential is often found in the physics literature, there is no physical interpretation for this specific functional form here. It serves solely as a convenient repulsive potential which has a continuous slope such its associated force can be calculated after discrete displacements.

This potential gives the point-like bead an effective diameter of σ. This same potential is used on geometrical constraints (eg., walls, cylinders...), in which case r is the distance normal to the surface to the bead centre. Care must be taken to define the effective size of objects since this potential effectively adds an exclusion length of $\sigma/2$ to the distance where the surface is nominally defined. For example, a two-sided mathematically thin wall will have an effective thickness of σ, corresponding to an exclusion layer of $\sigma/2$ per side.

Monomers are connected with Finitely-Extensible-Nonlinear-Elastic (FENE) springs [12, 14]

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}\, k_{\mathrm{FENE}}\, {r_0}^2 \ln\left[1 - \frac{r^2}{{r_0}^2}\right] \tag{23}$$

$$\tag{24}$$

where the maximum extension $r_0 = 1.5\sigma$ and a spring constant of $k_{\mathrm{FENE}} = 30\epsilon/\sigma^2$ is used. Similar to U_{WCA}, the form of this potential can be found in the physics literature. Taking the associated force $\mathbf{F}_{\mathrm{FENE}} = -\nabla U_{\mathrm{FENE}}$ will yield an approximation to the inverse of the previously mentioned Langevin function [7]. Recall that this would approximate the general force resulting between pulling two polymer ends. Again—although tempting—I do not wish to lend this interpretation to two given generic coarse-grained beads. It merely serves as a convenient form which has a finite extension (as opposed to a harmonic

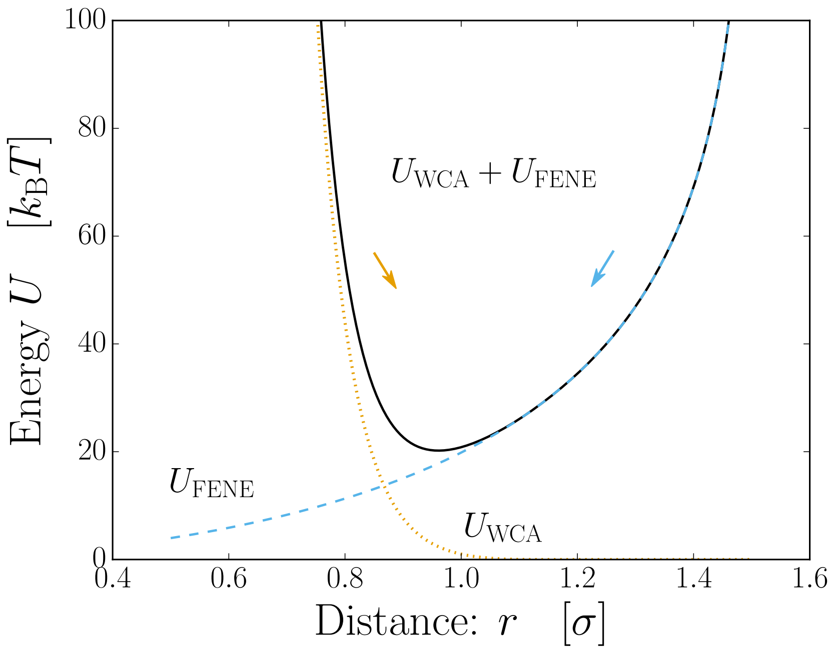


Figure 8: Combining the WCA with the FENE potential using the Kremer-Grest parameters yields a slightly asymmetric energy well surrounded by steep walls that prevent severe overlap and over stretching.

spring). Moreover, using the parameters $r_0 = 1.5\sigma$ and $k_\mathrm{FENE} = 30\epsilon/\sigma^2$ together with the repulsive U_WCA will give simulations where bond-crossing will be extremely unlikely [12, 14].

The two potentials are plotted in Fig. 8 as a function of the distance between two connected beads. The purely repulsive U_WCA combined with the purely attractive U_FENE gives an asymmetric well with the energy minimum at $r_\mathrm{min} \approx 0.9609\sigma$. Since the potential is not symmetrical, the mean value of the bond length has a temperature-dependence. At $k_\mathrm{B}T = 1.0\epsilon$ the mean is $\langle r \rangle \approx 0.9679\sigma$ whereas the lower value of $k_\mathrm{B}T = 0.02\epsilon$ brings a mean $\langle r \rangle \approx 0.9610\sigma$, closer to the well minimum value.

When the WCA length scale σ is considered as a bead diameter, bond length, or exclusion length from a surface, it is to be taken as *nominal* values—none are identical to unity. All of them are, however, close to unity which is the value I will be assuming in the following chapters (unless specified otherwise).

Furthermore, under the influence of a constant stretching force, both r_min and $\langle r \rangle$ will be displaced towards higher values. I will study some situations where this can add unnecessary complications. In Chapter 2, 5, and 6 the potentials U_WCA and U_FENE are rescaled by a multiplicative factor to reduce bond stretching, and thus avoid any such problems.

1.6 Modeling polymer translocation

Let us first consider the two steps: capture and translocation. The polymer capture process occurs at diffusive time scales that are much longer than the time scales for the translocation process. In a typical translocation experiment, the time relevant to these two processes differ by about two orders of magnitude [4]. Due to this separation of time scales, I will follow in the footsteps of the majority of coarse-grained simulation and theoretical work and initialize the polymer with at least one monomer inside the nanopore; the polymer end will be fixed inside the pore with the remaining monomers on the *cis*-side. The resulting polymer conformations will be essentially equivalent to that of a polymer tethered on a wall.

The polymer relaxation time can be interpreted as the time scale needed for a given polymer conformation to significantly change due to Brownian motion. In other words, it is the time needed to lose memory of the initial state. This time scale is often used interchangeably with the diffusion time τ_{diff}.

The relaxation time depends, amongst other things, upon the polymer size N, the friction coefficient ζ and the magnitude of thermal energy $k_{\text{B}}T$. The translocation time τ on the other hand depends upon the driving force F, N and ζ. The ratio of the relaxation time to the translocation time determines the driving regime. In Chapter 2, I will be investigating the N scaling between the two time scales and the amount of coarse-graining present in the simulation model. Experimental estimates for this time ratio will be presented and we will discuss how simulation parameters can be tuned in order to obtain a similar value.

Quasi-static translocation ($\tau_{\text{diff}} \ll \tau$) occurs where every step of the translocation process takes a time scale longer than the relaxation time τ_{diff}. In other words, the polymer sections on *cis* and *trans* have sufficient time to fluctuate and explore the phase-space during the progression of translocation.

In this limit the translocation process can be seen as a polymer crossing a free-energy barrier. The dynamics are quasi-static, and entropic costs can be estimated to map out the shape of the barrier. Some of the first theoretical descriptions of translocation provide predictions in this regime using a Fokker-Planck formalism [15, 16]. One must be careful to note that the driven translocation limit in this formalism still corresponds to situations of quasi-static translocations. Simulations by Kantor and Kardar failed to find the diffusive (or undriven) scaling prediction of $\tau \sim N^2$ [17]. Perhaps most importantly, they remarked that even in their undriven simulations, the quasi-static assumption was invalid, a result later confirmed by Steve Guillouzic [11] and Michel Gauthier, a PhD student in the Slater group [18].

Gauthier implemented the free-energy profiles obtained in the quasi-static regime as effective forces in a Monte-Carlo model [19, 20]. Due to the coarse-grained nature of the model, high N limits could be used to verify the various scaling behaviours predicted by the early theories. The $\tau \sim N^2$ scaling was later recovered with Langevin Dynamics in low viscosity conditions by Hendrick de Haan [21].

In the other extreme is the *highly-driven* translocation regime ($\tau_{\text{diff}} \gg \tau$). Here, the driving force deforms the polymer such that it does not have sufficient time to adapt before the translocation process proceeds to even further deformation.

This brings us to the theoretical contributions of Sakaue and co-workers [22, 23, 24, 25]. Their theoretical formalism borrows from previous ideas used to explain polymer stretching (deformation) under flow. The deformation of a tethered polymer under laminar flow falls into different regimes: *trumpet*, *stem-flower* and *stem*, depending on the ratio between the polymer relaxation time and the flow rate. Sakaue and co-workers applied these ideas to the situation of a polymer deformed due to a force applied to the monomer located in the pore. The out-of-equilibrium dynamics of the resulting polymer translocation are described with what is now called the *Tension-Propagation* theory.

The Ala-Nissala group implemented the forces predicted by the TP theory into a coarse-grained Brownian Dynamics model. This coarse-grained model was able to generate data for large N (up to 10^6) and found surprisingly long lasting finite-size effects [26]. The latter can reconcile discrepancies in N-scaling studies: the asymptotic values were simply not attained.

This is due to the effect of pore friction, which can contribute to a significant $t \sim AN^1$ term in the translocation time $\tau \sim AN^1 + BN^2$. The Polson group found that indeed, high friction situations are found to scale as N^1 in both quasistatic and driven translcocations [27]

In between these extremes lies a family of regimes which I will rarely be considering as my main focus here is the highly-driven regime. The latter regime corresponds to our experimental estimates of the experimental driving regime. Exceptions to this rule includes the simulations performed in Chapter 2 which are centred around how the modelling details can affect the ratio of time scales. In Chapter 5, I also investigate how different phenomena contribute to noise across different driving regimes.

The highly-driven regime can be pushed to a rather unphysical limit that I refer to as *deterministic* translocations. The lack of Brownian motion yields deterministic physics where $\tau_{\text{diff}} \to \infty$. This limit lends itself to simple theoretical calculations which can easily be compared to simulations with $k_{\text{B}}T = 0$. Although not practical, this provides valuable insight into translocation dynamics.

1.6.1 Tension-Propagation theory in the highly driven limit

As for the computer simulations, the theoretical model presented herein will consider a polymer initialized in half-space, with a monomer inside the pore. The driving field is modelled as an applied force on the monomer inside the nanopore. We assume here that monomers travel in a single-file sequential manner to the *trans*-side.

It is best to envision the theoretical Tension-Propagation model in the highly-driven limit as a purely deterministic process. Figure 9 depicts snapshots at different times obtained by deterministic simulations with $k_{\text{B}}T = 0$. The initial conformation is overlaid on to all the frames as black circles. Cleary, as translocation progresses from frame a) to c), the blue monomers far from the nanopore which are not affected by the field do not move from their initial $t = 0$ positions. The monomers colored in orange move under the influence of a force applied on the monomer inside the pore only. The force propagates down the chain via connected monomers. The translocation process is divided into two main stages; Tension-Propagation (Fig. 9 a-c); and Post-Propagation (Fig. 9 d-f).

In Fig. 9 two monomers are labeled k and s. The monomer in the pore is labeled s. At the beginning of the process $s = 0$ (Fig. 9a) and at the end of the process, $s = N = 100$ (the time frame after Fig. 9f) [3]. This will be used as a *translocation coordinate* and helps describe the overall progression of translocation.

The monomer labeled k serves to indicate the tension front. Similar to the translocation coordinate, at the beginning $k = 0$. The tension front propagates down the chain and up to the last monomer $k = N = 100$ during the Tension-Propagation phase (Fig. 9a-c), after which it remains at $k = 100$ for the remaining Post-Propagation (Fig. 9d-f).

During all times the effect of the monomers on the *trans*-side will be neglected. They are thus masked from view in Fig. 9.

Consider how the first monomer in the pore, $s = 0$, reacts to the application of a driving force of magnitude F oriented in the x-direction towards the *trans*-side. In the

---

[3] Note here that I use again the approximation $N - 1 \approx N$ for simplicity.

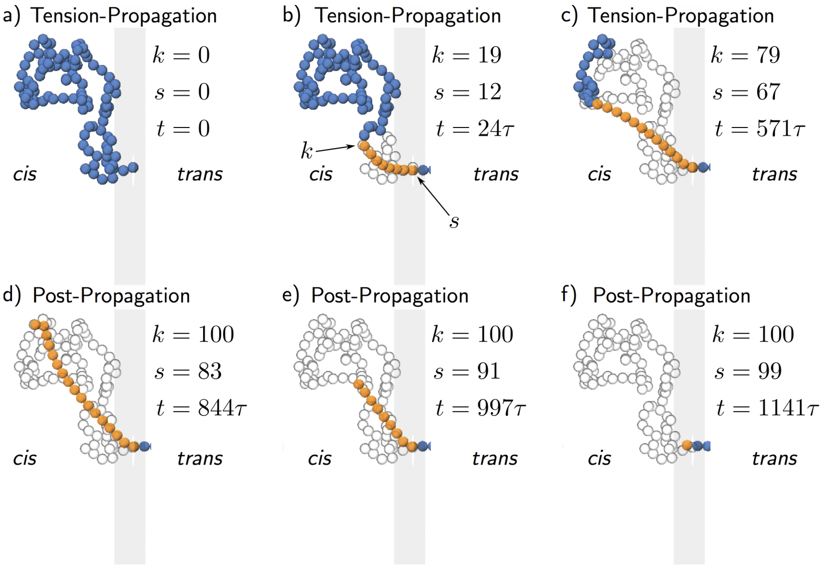


Figure 9: Simulation snapshots (at $k_{\mathrm{B}}T = 0$) depicting steps of translocation process according to Tension-Propagation theory for a $N = 100$ polymer. Monomers on the *trans*-side are hidden from view. The two stages of translocation are the propagation of tension a)–c) and the remaining post-propagation. Monomers under tension are coloured in orange. The initial (at $t = 0$) polymer conformation is shown as monomer outlines.

overdamped limit, this force brings movement of terminal velocity $v = F/\zeta$. However, this motion is short-lived because the connected next monomer feels a pull as the tension propagates down the chain. The first monomer, upon sufficient displacement, will exit the pore and will no longer be driven by the applied force. If we neglect the monomers on the *trans*-side, the force balance equation is generalized to

$$F = (k - s)\,\zeta v, \tag{25}$$

where the $(k - s)$ monomers set in motion by the applied force each provide a drag force ζv. The above equation provides the *translocation rate*

$$\frac{ds}{dt} = \frac{F}{\zeta b}\frac{1}{(k-s)}.$$

This suggests that the fundamental time unit here is $\zeta b/F$ which can be seen as a characteristic time for a driven displacement of a single bead at terminal velocity F/ζ. The term $(k - s)$ simply reflects the number of monomers dragged by the driving force. Inspection of Fig. 9 shows how one can approximate the number of dragged monomers by the spatial distribution of monomers present in the initial conformation. Consider the position of monomer 19 from Fig. 9b. Note how its position is the same as in the initial frame Fig. 9a. Furthermore, note how the $(k-s)$ coloured monomers form a relatively straight segment between monomer k and monomer s. Since monomer s is always in the nanopore, the distance between monomer k and the nanopore can be written as

$$R_k \cong (k-s)\,b \tag{26}$$

where b is the monomer diameter. Note that the set of distances $\{R_k\}$ for $k \in (0, N)$ is completely captured in the initial conformation. This also provides a way to find the translocation coordinate at the end of the Tension-Propagation phase

$$s^* = N - R_N/b. \tag{27}$$

For simplicity, we will use the dimensionless distances $\tilde{R}_k = R_k/b$.

The time derivative of Eq. 26 gives

$$\frac{d\tilde{R}_k}{dt} = \frac{dk}{dt} - \frac{ds}{dt},$$

which, when combined with Eq. 25, gives

$$\frac{F}{\zeta b} = \tilde{R}_k \left(\frac{dk}{dt} - \frac{d\tilde{R}_k}{dt} \right).$$

By separation of variables, the temporal dt term is brought to the LHS to give

$$\frac{F}{\zeta b} dt = \tilde{R}_k dk - \tilde{R}_k d\tilde{R}_k.$$

The Tension-Propagation time τ_{TP} is found by integrating over the intervals $k \in (0, N)$ and $\tilde{R}_k \in (0, \tilde{R}_N)$

$$\int_0^{\tau_{\text{TP}}} \frac{F}{\zeta b} dt = \int_0^N \tilde{R}_k dk - \int_0^{R_N} \tilde{R}_k d\tilde{R}_k$$

$$\boxed{\tau_{\text{TP}} \frac{F}{\zeta b} = \int_0^N \tilde{R}_k dk - \frac{1}{2}\tilde{R}_N^2} \tag{28}$$

Thus, the time τ_{TP} required for the tension front to reach the last monomer (N), depends upon the conformation via the list of distances R_k of all the monomers to the pore. The translocation coordinate $s^* \equiv s(t = \tau_{\text{TP}})$ at this instant is given by Eq. 27.

We now turn to the process where the tension front remains at the last monomer $k = N$ and the remaining $(N-s) = \tilde{R}_N$ monomers are driven through the pore. In this Post-Propagation phase, $dk/dt = 0$ and one can write the force-balance relation as

$$\frac{F}{\zeta b} = \tilde{R}_N \frac{ds}{dt},$$

where we note that although k does not evolve, the scaled distance $\tilde{R}_N$ now *decreases* with time. The variables can be separated to give

$$\frac{F}{\zeta b}dt = Nds - sds.$$

It is straightforward to integrate this equation over time and the translocation coordinate $s \in (s^*, N)$:

$$\int_0^{\tau_{\mathrm{PP}}} \frac{F}{\zeta b}dt = \int_{s^*}^{N} Nds - \int_{s^*}^{N} sds.$$

The integrals are easily evaluated to obtain

$$\tau_{\mathrm{PP}}\frac{F}{\zeta b} = N^2 - Ns^* + \frac{1}{2}s^{*2} - \frac{1}{2}N^2,$$

which one can simplify using Eq. 27 to obtain the Post-Propagation time

$$\boxed{\tau_{\mathrm{PP}}\frac{F}{\zeta b} = \frac{1}{2}\tilde{R}_N^2}. \tag{29}$$

Adding the time needed to complete each of the two phases from Eqs. 28 and 29 gives the complete translocation time of the purely deterministic process:

$$\boxed{\tau = \tau_{\mathrm{TP}} + \tau_{\mathrm{PP}} = \int_0^N \tilde{R}_k dk}. \tag{30}$$

The input for Eq. 30 is a function (list) of initial monomer distances $\{R_k\}$ which needs to be integrated (summed) for a continuous (discrete) model. This can be evaluated via a model for R_k or results from simulations. In different chapters of this thesis, the translocation theory can be summarized as the simple evaluation of $\tau(\{R_k\})$, for different initial conditions. In particular, I will be focusing on how $\{R_k\}$ is affected by pre-confinement geometries (Chapters 3 and 4), the application of stretching forces (Chapter 6) and polymer stiffness (Chapter 8).

In Chapter 3, I investigate the effect of an additional term to account for pore friction and monomer crowding on the *trans*-side, both of which were neglected in the preceding derivation. Monomer crowding on the *trans*-side is explicit in the theoretical derivation of Chapter 8 since polymer rigidity increases the correlation length which can carry over to the *trans*-side.

In the conclusion, I discuss possible different directions this research can take. As is often the case with scientific research, the attempt to answer one question ends with asking three.

For an additional perspective before proceeding to the next chapters, at this point the reader may want to read Appendix A. One of the systems of study is presented in a language suitable for a wider audience. This should give the general flavour of the remaining chapters in a more accessible form and can serve as a small introductory supplement.

2

Using a Péclet number for the translocation of a polymer through a nanopore to tune coarse-grained simulations to experimental conditions

Hendrick W. de Haan, **David Sean**, Gary W. Slater

This paper came to be as the result of multiple discussions between Hendrick de Haan and I. Hendrick had the original idea of using a Péclet number, he generated the simulations as well as provided the general direction of the paper. I took charge of the literature review, generated the plots with Hendrick's data, provided Tension-Propagation related insight, and contributed to approximately 20% of the text. I have Hendrick's explicit permission to use this paper.

Published in *Physical Review E*
Reprinted with permission from [Hendrick W. de Haan, David Sean, and Gary W. Slater, Physical Review E, **91**, 022601 (2015)].

http://doi.org/10.1103/PhysRevE.91.022601

PHYSICAL REVIEW E **91**, 022601 (2015)

Using a Péclet number for the translocation of a polymer through a nanopore to tune coarse-grained simulations to experimental conditions

Hendrick W. de Haan
Faculty of Science, University of Ontario Institute of Technology, Oshawa, Ontario L1H 7K4, Canada

David Sean and Gary W. Slater
Physics Department, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

(Received 7 July 2014; published 12 February 2015)

Coarse-grained simulations are often employed to study the translocation of DNA through a nanopore. The majority of these studies investigate the translocation process in a relatively generic sense and do not endeavor to match any particular set of experimental conditions. In this manuscript, we use the concept of a Péclet number for translocation, P_t, to compare the drift-diffusion balance in a typical experiment vs a typical simulation. We find that the standard coarse-grained approach overestimates diffusion effects by anywhere from a factor of 5 to 50 compared to experimental conditions using double stranded DNA (dsDNA). By defining a Péclet control parameter, λ, we are able to correct this and tune the simulations to replicate the experimental P_t (for dsDNA and other scenarios). To show the effect that a particular P_t can have on the dynamics of translocation, we perform simulations across a wide range of P_t values for two different types of driving forces: a force applied in the pore and a pulling force applied to the end of the polymer. As P_t brings the system from a diffusion dominated to a drift dominated regime, a variety of effects are observed including a non-monotonic dependence of the translocation time τ on P_t and a steep rise in the probability of translocating. Comparing the two force cases illustrates the impact of the crowding effects that occur on the *trans* side: a non-monotonic dependence of the width of the τ distributions is obtained for the in-pore force but not for the pulling force.

DOI: 10.1103/PhysRevE.91.022601 PACS number(s): 82.35.Lr, 87.15.ap, 82.35.Pq

I. INTRODUCTION

The translocation of a polymer through a nanopore has been the subject of intense study in recent years, both due to its biological relevance as well as emerging nanotechnology applications such as nanopore-based devices for detecting and sequencing DNA. Among this body of work, there have been numerous molecular dynamics simulation studies employing a coarse-grained (CG) approach in which a relatively generic polymer is simulated. As shown in Table I, a typical CG setup consists of a polymer on the order of 100 monomers ($N = 100$), an external driving force of $F = 1$–$10\,\epsilon/\sigma$ (where σ is the monomer or bead diameter, and ϵ is the characteristic energy, typically the strength of the bead-bead interaction), and a thermal energy of $k_BT \approx \epsilon$. In this paper, we consider the physical appropriateness of such setups for modeling the translocation of polyelectrolytes such as double stranded DNA (dsDNA). We thus consider a typical setup of $N = 100$, $F \approx \epsilon/\sigma$, $k_BT \approx \epsilon$ and investigate what experimental conditions this corresponds to.

To do so, we define a Péclet number for translocation, P_t, such that we can quantify the balance between the drift (external force) and diffusive (thermal noise) components of the translocation process. We then introduce a factor λ which we use as a control parameter for the translocation Péclet number. We show that adjusting the temperature variable k_BT in the Langevin dynamics (LD) thermostat, which is inversely proportional to λ, is an appropriate and convenient method of implementing λ into the simulations. Performing simulations across a range of λ values then permits the isolation of the effects that the drift-diffusion balance has on the translocation process. Low λ values correspond to a diffusion dominated process while high λ values yield a drift dominated process.

To test the role of the diffusion-drift balance, we simulate polymer translocation across a wide range of P_t values. We demonstrate that the translocation process changes significantly as we move from a diffusive to a drift dominated regime. Effects include a steep rise in the probability of translocation, a non-monotonic increase in the translocation time, and a non-monotonic decrease in the spread of translocation times.

At the end of the paper, we consider different experimental translocation scenarios: dsDNA, ssDNA, and rod-like viruses. These examples demonstrate the usefulness of using P_t to match simulations to experimental conditions via measurable quantities. By estimating the λ value appropriate for each scenario, we also show that the standard approach is unphysical for certain cases. Most dramatically, we find that typical CG setups for dsDNA overestimate diffusive effects by a factor of 5–50.

II. THEORY

A. Péclet number for translocation

As for most transport-based processes, the translocation of a polymer through a nanopore contains aspects of both diffusive and driven dynamics, the latter arising primarily from the presence of external forces. The balance between these two mechanisms can greatly affect the dynamics of translocation. In the limit of zero field, or unbiased translocation, a polymer started halfway through the pore will exhibit stochastic, random-walk behavior (albeit subdiffusive) for the majority of the process. On the other hand, in the limit of zero diffusivity, the trajectory of a given translocating polymer will be deterministic.

TABLE I. Values of the degree of polymerization N, the driving force F, and the thermal energy k_BT found in the literature for three-dimensional Langevin dynamics simulations of translocation (* indicates estimated values).

Reference	N	F (units of $\frac{\epsilon}{\sigma}$)	k_BT (ϵ)
[1]	100–500	1–20	1.2
[2]	100–400	10	1
[3]	256	0.05–5	1.2
[4]	200	0–2	1
[5]	16–256*	0.5–10	1.2
[6]	100	0–10	1
[7]	8–256	4–6	1.5
[8]	31–251	0.3–10	1
[9]	100	1–100	1

To characterize the balance between drift and diffusion, we define a dimensionless parameter given by the ratio of the free solution polymer relaxation time τ_{relax} over the translocation time τ_{trans}. As it characterizes the drift-diffusion balance, this metric is essentially a Péclet number [10] for translocation:

$$P_t = \frac{\tau_{\text{relax}}}{\tau_{\text{trans}}}. \tag{1}$$

A similar argument was given by Saito and Sakaue [11] as well as Dubbeldam *et al.* [6] where a Péclet number was given in terms of simulation parameters such as the friction coefficient. Our goal here is to match coarse-grained simulations to experimental conditions and thus we define P_t in terms of quantities obtainable both experimentally and *in silico*. As usual, the free solution-relaxation time is given by the expression

$$\tau_{\text{relax}} = \frac{R_{\text{go}}^2}{D_o}, \tag{2}$$

where R_{go} is the equilibrium radius of gyration and D_o is the free solution diffusion coefficient. Our Péclet number is then given in terms of three observables:

$$P_t = \frac{1}{\tau_{\text{trans}}}\frac{R_{\text{go}}^2}{D_o}. \tag{3}$$

In order to match experimental and simulation results, we will now develop quantitative expressions for P_t for both cases. Starting with the latter, it is first necessary to introduce our simulation approach.

B. Simulation setup

1. *Langevin dynamics simulations*

We employ a coarse-grained simulation approach that is very common in molecular-dynamics based translocation work [1–9,12]. Performing Langevin dynamics (LD), the effects of the solvent are included implicitly and the resulting equation of motion is

$$m\dot{\vec{v}} = \vec{F} - \vec{\nabla}U(\vec{r}) - \gamma\vec{v} + \sqrt{2\gamma k_BT}\vec{\xi}(t), \tag{4}$$

where m is the mass of the particle, $\vec{v}$ is the velocity, $\vec{F}$ is the applied external force, $-\vec{\nabla}U(\vec{r})$ is the sum of the conservative forces, γ is the friction coefficient, and the term $-\gamma\vec{v}$ represents the damping effects of the fluid. The last term in Eq. (4), in which $\vec{\xi}$ [12] is a random vector, models the random kicks of the solvent.

From Eq. (4), it is straightforward to show that for a free particle ($\nabla U(\vec{r}) = 0$), the diffusion coefficient D and drift velocity v_{drift} are given by

$$D = \frac{k_BT}{\gamma} \tag{5}$$

and

$$\vec{v}_{\text{drift}} = \frac{\vec{F}}{\gamma}. \tag{6}$$

Finally, using the equipartition theorem, the thermal velocity of the particle is given by

$$v_{\text{th}} = \sqrt{\frac{3k_BT}{m}}. \tag{7}$$

2. *Coarse-graining*

Equations (5), (6), and (7) characterize the dynamics of a single particle. In a typical coarse-graining procedure, we take a single simulation bead as a model for many individual particles of mass m. Consider such a bead to represent λ smaller units. In the case of free-draining hydrodynamics, each particle contributes independently to the friction of the simulation bead such that its net friction is $\lambda\gamma$. Similarly, the mass of the bead is λm, and if each subparticle feels a force F, the net force is λF. Hence, using the transformations

$$\gamma \to \lambda\gamma, \tag{8}$$

$$\vec{F} \to \lambda\vec{F}, \tag{9}$$

$$m \to \lambda m, \tag{10}$$

Eqs. (5), (6), and (7) become

$$D = \frac{k_BT}{\lambda\gamma}, \tag{11}$$

$$\vec{v}_{\text{drift}} = \frac{\lambda\vec{F}}{\lambda\gamma} = \frac{\vec{F}}{\gamma}, \tag{12}$$

$$v_{\text{th}} = \sqrt{\frac{3k_BT}{\lambda m}}. \tag{13}$$

The drift velocity is thus unaffected in the free-draining limit. We will use λ as a Péclet control parameter (with this general definition, it is possible to use $\lambda < 1.0$). As will be shown, a specific λ must be implemented in order to achieve the correct, experimentally relevant Péclet number in the simulations. Moreover, for a given setup, performing simulations across a range of λ values allows us to assess the impact that the drift-diffusion balance has on the translocation process. In this work, we take a standard setup and show that the results such as the probability of translocation and the net translocation time depend significantly on λ. To conclude, it is shown that for simulations modeling the translocation of dsDNA, typical simulations setups significantly overestimate diffusion effects.

3. Polymer simulations

To model the forced translocation of a polymer chain, we use a common CG approach [12]. A purely repulsive WCA potential is used for excluded volume interactions between beads [13]:

$$U_{\mathrm{WCA}}(r) = \begin{cases} 8\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] & \text{for } r < 2^{\frac{1}{6}}\sigma, \\ 0 & \text{for } r \geqslant 2^{\frac{1}{6}}\sigma. \end{cases} \qquad (14)$$

This defines the simulation units of length (σ, the size of the monomers) and energy (ϵ). The Weeks-Chandler-Andersen (WCA) potential is also used to model the membrane as a continuous surface with a pore of radius 1σ. This ensures single file translocation [14]. To link connected monomers, the finitely extensible nonlinear elastic (FENE) potential is used; similar to the work of Kremer and Grest, we set its maximum spring extension to 1.5σ [12,15]. For reasons which will be clarified later, stable interactions are needed between connected beads over an unconventionally wide range of thermal forces. To achieve this, we increased the prefactors for the WCA and FENE potentials by a factor of 2 from what is typically found. We thus use a FENE spring constant of $k = 60\,\epsilon/\sigma^2$.

As the probability of translocation is very low at low λ values, we initiate the polymer with one monomer on the *trans* side. We examine two different force scenarios: a driving force $F = 1.0\,\epsilon/\sigma$ applied to the monomer(s) located in the pore, and a pulling force $F = 1.0\,\epsilon/\sigma$ applied on the end monomer (see insets of Fig. 3). We use Eq. (12) to derive a simulation time unit independent of both λ and $k_{\mathrm{B}}T$ as $\tau_{\mathrm{sim}} = \sigma/v_{\mathrm{drift}} = \sigma\gamma/F$. From now on, we will drop the LD units for simplicity.

C. The simulation Péclet number

From the definition in Eq. (3), we require the free solution diffusion coefficient, the equilibrium radius of gyration, and the translocation time of a polymer of N beads. Since there is no hydrodynamic coupling in LD, the net friction coefficient scales like $D_o = k_{\mathrm{B}}T/\gamma N$. Incorporating the Péclet control parameter λ we obtain

$$D_o = \frac{k_{\mathrm{B}}T}{\lambda\gamma N}. \qquad (15)$$

Simulations were performed at $k_{\mathrm{B}}T = \epsilon$ to determine τ_{trans} as a function of F, N, and γ. Our data agree with

$$\tau_{\mathrm{trans}} = A\frac{\gamma}{F}N^{\alpha}. \qquad (16)$$

For $F = 1.0$, $\gamma = 1.0$, and $N = 50,100,200$, $A = 3.28\sigma$ and $\alpha = 1.43$ is the effective scaling exponent, in good agreement with previous studies (e.g., [5,7]). These parameters are only weakly dependent on $k_{\mathrm{B}}T$ (data not shown). Finally, the equilibrium radius of gyration can be fit using the standard expression

$$R_{\mathrm{go}} = BN^{\frac{3}{5}}, \qquad (17)$$

where $B = 0.468\sigma$ in our case. Using Eq. (3), our simulation translocation Péclet number is thus given by

$$P_{\mathrm{sim}} \approxeq \left(\frac{F}{A\gamma N^{\alpha}}\right)\frac{\left(BN^{\frac{3}{5}}\right)^2}{\frac{k_{\mathrm{B}}T}{\lambda\gamma N}} \qquad (18)$$

$$\approxeq \frac{C\lambda F N^{0.77}}{k_{\mathrm{B}}T}, \qquad (19)$$

where $C = 0.067\sigma$ (P_{sim} is dimensionless as required). Note that P_{sim} is independent of the friction coefficient.

The scaling $P_{\mathrm{sim}} \sim \lambda$ allows us to use λ as a *control parameter* which changes the Péclet number. In practice though, a LD simulation does not directly include λ; instead, λ is used in an implicit way when interpreting the simulation data. However, Eq. (19) suggests that we can effectively tune λ (as we shall see, we will need to increase P_{sim}) by changing one (or any combination of) the three simulation parameters N, F, or $k_{\mathrm{B}}T$ as discussed below.

(1) Using longer polymers by increasing N will increase P_{sim}, but this quickly becomes prohibitively expensive in terms of computation time.

(2) While increasing F works well for small increases in F, it generally becomes problematic at very large forces because the friction parameter is unchanged. Unless one is careful, very high forces may lead to situations where we are no longer in the overdamped regime that is intrinsic to nanofluidics. Additionally, Eqs. (11)–(13) show that increasing F is not fully equivalent to increasing λ; in other words, increasing F actually changes the nature of the problem. Furthermore, one would have to take into account the fact that higher forces will also yield faster translocations (note also that τ_{sim} varies with F). This effect would need to be corrected for in order to study the dependence upon the Péclet number only.

(3) We can also change P_{sim} by varying the thermal factor $k_{\mathrm{B}}T$ in the simulations. This is clearly the best approach since Eqs. (11)–(13) indicate that what actually matters is the ratio $k_{\mathrm{B}}T/\lambda$. Changing $1/k_{\mathrm{B}}T$ is thus fully equivalent to changing λ.

In order to reduce computation costs and to ensure that we remain in the overdamped regime we will follow the third prescription. Hence, changes in λ are implemented by setting the temperature of the simulation thermostat to

$$k_{\mathrm{B}}T = \frac{1}{\lambda}[\epsilon]. \qquad (20)$$

Note that as the external force F, the bead friction γ, and the energy unit ϵ are kept fixed throughout, the simulation time unit $\tau_{\mathrm{sim}} = \sigma\gamma/F$ remains constant when λ is changed. Therefore, the correspondence between the simulation molecular weight or our freely jointed polymer and that of a real polymer will be determined solely by the choice of N and the length scale σ.

III. RESULTS

To demonstrate how the Péclet number can affect translocation dynamics, we have performed simulations of driven translocation across a wide range of λ. Simulations were performed for polymers of length $N = 50$ and $N = 100$. We focus on the $N = 50$ results as it is easier to resolve all relevant regimes for the shorter polymer. Two driven translocation cases were studied: (i) a force applied to any monomer that is in the pore and (ii) a force applied to the lead monomer which pulls the polymer through the pore [16]. The results for the translocation time are shown in Fig. 1. In discussing the results,

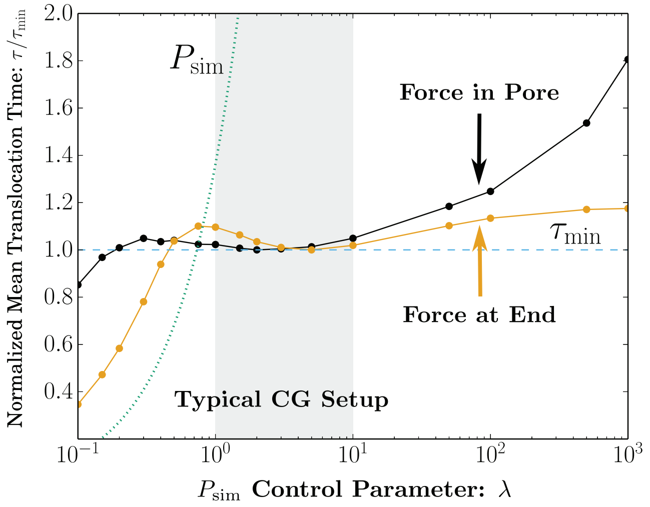


FIG. 1. (Color online) Translocation time τ (normalized by the local minimum $\tau_{\rm min}$) vs λ for the in-pore force case (black) and for pulling at the end case (orange) for a polymer chain of size $N = 50$. The corresponding simulation Péclet number $P_{\rm sim}$ is shown as a dotted green line.

we present the data in terms of the dependence on the Péclet control parameter, λ, where different values of λ correspond to different Péclet numbers.

Non-monotonic results are obtained for both force scenarios. From the results, we define three regimes. At low λ values (high T), τ increases with increasing λ. For intermediate λ values, τ decreases slightly with increasing λ. And finally, at high λ values (low T), τ again increases with increasing λ. Given the definition of P_t, we can associate these three regimes with the relative drift-diffusion balance: (i) low λ values correspond to a diffusion dominated regime, (ii) intermediate λ values correspond to a transition regime, and (iii) high λ values correspond to a primarily driven regime.

These regimes can be verified by examining the probability of translocation as shown in Fig. 2. At low λ, diffusion is dominant over drift and thus the polymer frequently retracts to the *cis* side. At high λ, the thermal noise is suppressed and thus, as the process becomes purely deterministic, the polymer never retracts and the translocation probability goes to 1. In the transition regime, the probability rapidly increases between these two limits (note the x axis is logarithmic). Having outlined three regimes, we explore the physics of translocation in each one.

A. Diffusive: Low λ

At low λ, τ increases with increasing λ. This regime corresponds to the quasi-static limit for translocation. Given the high diffusivity, the relaxation time of the polymer is very short and in fact is shorter than the typical time for monomers to translocate through the pore. Hence, the polymer is relaxed at each stage of translocation. This allows for an approximation in which the polymer is reduced to a single particle between two absorbing walls subject to drift and diffusion while crossing an entropic barrier [17–19]. The validity of this approximation has been demonstrated in simulations where the high diffusivity is achieved by lowering the viscosity of the fluid [20,21].

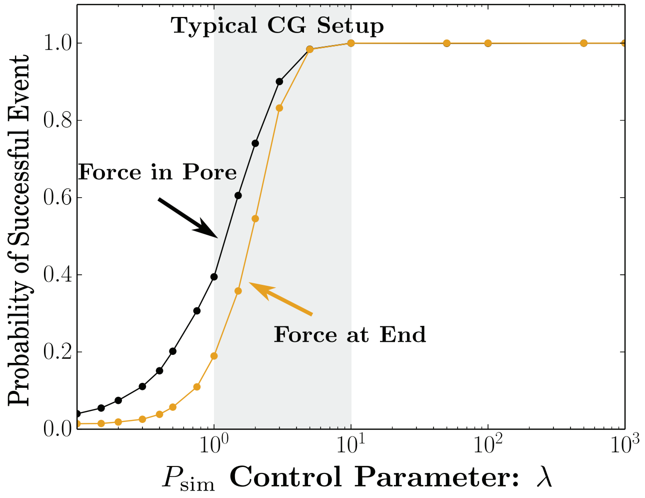


FIG. 2. (Color online) Probability of a successful translocation event vs λ for the in-pore force (black) and force at end (orange) cases for a polymer chain of size $N = 50$.

In this regime, two factors contribute to τ decreasing as λ decreases. First, a higher diffusion coefficient means that the rate of motion for the polymer increases which result in faster translocation, and thus higher diffusion aids translocation. More importantly, there is also a selection process implicit to the dynamics at low λ. This can be seen in Fig. 2 where the probability of translocation is very low for $\lambda < 1.0$. As λ decreases, the probability of translocation decreases further. This means that translocation only occurs for those events in which the polymer quickly moves towards the *trans* side; i.e., only the fastest events survive and τ decreases.

To examine the details of how τ increases with λ, the distributions for four selected λ values are shown in Fig. 3. For both the in-pore and pulling forces, at low λ values, the distributions shift to longer translocation times as λ increases; this agrees with the physical picture outlined above.

B. Transition: Intermediate λ

Examining the distributions beyond the low λ values, the nature of the transition region becomes clear. Although the distributions continue to shift to larger τ values, the width of the distributions begins to decrease significantly. As the dynamics transition from diffusion to drift, the contribution to the variance in τ arising from the thermal noise rapidly diminishes. Most significantly, this reduction in the impact of the stochastic path reduces the instances of long-time events. Hence, even though the most probable value of τ is increasing, the mean of τ is actually decreasing. This yields the non-monotonic nature of the transition region.

To quantify this, the standard deviation, σ_{st}, normalized by the mean translocation time τ for all λ values is shown in Fig. 4. For the pulling force, the normalized width continuously decreases with increasing λ. For the in-pore force case, σ_{st}/τ decays with increasing λ both in the diffusion and transition regimes, but it increases when λ is very large.

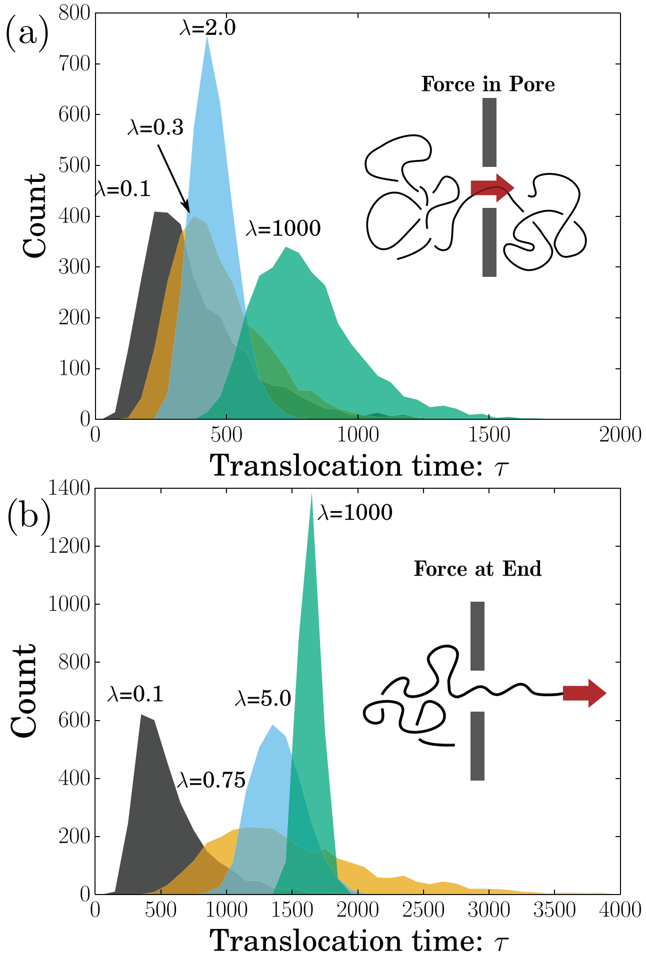


FIG. 3. (Color online) Distribution of translocation times for the force in pore (top) and force at end (bottom) cases for a polymer chain of size $N = 50$. In both cases, four different λ values are shown.

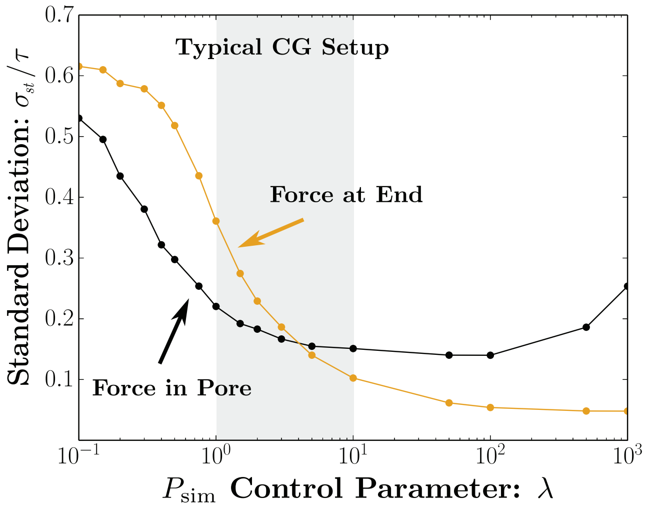


FIG. 4. (Color online) Standard deviation of the translocation time distributions, σ_{st} (normalized by the mean, τ) vs λ for a polymer of size $N = 50$.

C. High λ: Driven translocation

For the pulling force case, the normalized standard deviation σ_{st}/τ at high λ values appears to be plateauing at a finite value. As shown in Fig. 1, τ plateaus at high λ values. If thermal factors are negligible, the dynamics are essentially fully deterministic, and hence increasing λ has no further effect. Recall there are two factors contributing to the variance in τ for the pulling case: variations in the stochastic path and in the initial conformation of the polymer [11,22]. As λ increases, the first factor becomes negligible but the second remains since it is independent of the λ value. Hence, both τ and σ_{st} level off at high λ and thus so does the normalized width σ_{st}/τ.

The behavior is very different for the in-pore force where σ_{st}/τ increases with λ at high λ (this increase can also be seen in the distribution plots, Fig. 3). This difference in the σ_{st}/τ plots highlights one of the main differences between the two force cases: for the pulling force, there is no crowding on the *trans* side. For the in-pore force, once the monomers are through the pore, there is no force on them. At high λ values, they do not diffuse away quickly and thus significant crowding is observed. This crowding presents a steric barrier to translocation: the translocation time for the in-pore force increases much more dramatically than for the pulling force where it is actually levelling off (Fig. 1).

As τ increases, one expects σ_{st} to increase as well since the stochastic path will have a larger impact. However, σ_{st}/τ increases at high λ indicating that σ_{st} is increasing at a faster rate than τ. The reason for this is that crowding also presents an additional source of variation for the translocation time. The degree of crowding for any particular event will depend on the details of individual *trans* conformations. Hence, the translocation time is dependent not only on the initial conformation and the stochastic path, but also the *trans* conformation. This extra source of variance results in σ_{st}/τ increasing at high λ. Again, this is dramatically different from the pulling case where only the first two factors are present and σ_{st}/τ levels off.

Considering the increase in τ in this high λ regime, recall that the probability of translocation is ≈1.0. Hence, the increase in τ with increasing λ cannot be due to a selection process as it was in the low λ, diffusive regime. The increase can be seen by considering the relaxation of a *cis* subchain that just lost monomers to translocation: the relaxed conformation of this shorter subchain has a center-of-mass closer to the pore. Hence relaxation on the *cis* side biases the remaining monomers towards the pore, aiding translocation. As λ increases, this relaxation is suppressed (the *cis* subchain remains stretched away from the pore) and thus τ increases. A similar effect occurs on the *trans* side where instead, the relaxation carries the monomers away form the pore. Suppressing this relaxation increases monomer crowding on the *trans* side. Comparing the τ increase between the end-pulling (no crowding) and the in-pore force (crowding) highlights how the Péclet number can affect *trans* crowding.

IV. EXPERIMENTAL SCENARIOS: TUNING λ

In this section we examine several different experimental translocation scenarios. For each case, we first estimate the experimental value of the Péclet number and we then find how λ can be used to tune the simulations appropriately.

A. Double stranded DNA

Smith *et al.* [23] measured the diffusion coefficient for a wide range (4–309 kbp) of dsDNA sizes. From their data we get

$$D_o \approx \frac{2.38}{L^{0.608}}\ \mu\text{m}^2/\text{s}, \tag{21}$$

where L is the polymer length in μm. Smith *et al.* also extracted the radius of gyration from their data, from which we obtain

$$R_{\text{go}} = 0.146 L^{\frac{3}{5}}\ \mu\text{m}. \tag{22}$$

For the translocation time, we use two seminal studies for dsDNA through solid state pores. First, Storm *et al.* in 2001 studied the translocation of DNA strands 2.2–32.6 μm in length and found a scaling of $\tau \sim L^{1.27}$ [24]. From their data we estimate

$$\tau_{\text{trans}} \approx 5.6 \times 10^{-5} L^{1.27}\ \text{s} \tag{23}$$

However, Chen *et al.* obtained a linear scaling for DNA 1.0–16.5 μm in length [25]:

$$\tau_{\text{trans}} \approx 1.0 \times 10^{-4} L^{1}\ \text{s} \tag{24}$$

Both of these results are for an applied voltage of 120 mV, which is a typical value [26–29]. Voltages as low as 20 mV [30] and as high as >800 mV [25] have also been used. Using the τ_{trans} forms given above, we obtain

$$P_{\text{Storm}} \approx 160 L^{0.54}, \tag{25}$$

$$P_{\text{Chen}} \approx 90 L^{0.81}. \tag{26}$$

The scaling exponents (for N or L) found in the expressions for P_{sim}, P_{Chen}, and P_{Storm} are slightly different. One significant deviation between the simulations and experiments is that the simulations do not include hydrodynamic effects. In spite of these differences, the scaling is rather similar and thus achieving the correct P_t at one polymer size will result in essentially the correct P_t for a fairly wide range of sizes, as we shall see.

The last step before we can compare simulation and experimental Péclet numbers is the selection of a length scale, i.e., we need to choose the monomer bead size (σ). For dsDNA we can set the bead diameter to be 10 nm, a convenient value that is only slightly larger than estimates for the effective width of DNA (5–10 nm) [31–33]. Since the simulation pore radius is set to be σ, $\sigma = 10$ nm implies a pore with a diameter of 20 nm. This compares well to the values of 10 nm by Storm *et al.* [24] and 15 nm in Chen *et al.* [25]. This choice also yields a reasonable value of 10 nm for the effective width of dsDNA. However, it would require a persistence length of 5σ in order to properly model dsDNA—a feature that is neglected in the freely jointed chain (FJC) model that is often used in computational studies of polymer translocation (the effects of semiflexibility have been recently explored [34]).

With $\sigma = 10$ nm, a polymer of 100 beads corresponds to a DNA fragment with a contour length of 1 μm. This corresponds to the lower bound of the Chen data cited above, but if we consider doing a scaling experiment and extending N up to 400, then the corresponding lengths would encompass the lowest two or three data points of both the Chen and Storm data.

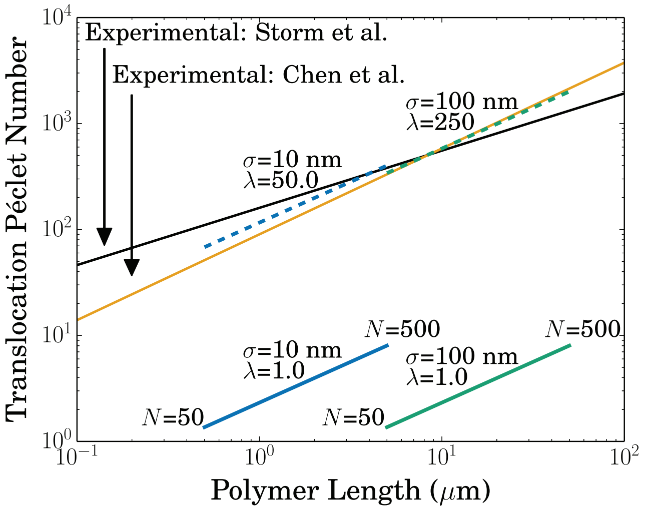


FIG. 5. (Color online) Experimental and simulation Péclet numbers as a function of polymer length. The experimental P_t based on the results of Storm *et al.* [24] and Chen *et al.* [25] have slightly different scalings with respect to polymer length. The simulation P_t are shown for two different length scales: $\sigma = 10$ nm is shown in blue and $\sigma = 100$ nm is shown in green. For the former, $\lambda = 50$ brings the simulation and experimental P_t into good agreement; for the latter, $\lambda = 100$ is required.

The Péclet numbers from Eqs. (19) and (26) are shown in Fig. 5. For $\sigma = 10$ nm and $\lambda = 1.0$, P_{sim} is significantly lower than the estimate from experiments for a given contour length. Hence, this CG model yields an artificially low Péclet number: diffusion is much greater than it should be. To correct this, we can choose $\lambda = 50$: this brings the data for P_t from simulation and experiments into much closer agreement. This indicates that to approximate the experimental balance between diffusion and drift, the diffusivity of the polymer should be lowered to 1/50 of its "default" simulation value.

To compare these values to typical CG setups, refer again to the list of parameters used in various simulation studies as shown in Table I. Most studies take $k_{\text{B}}T \gtrsim \epsilon$, a force value of 1–10 ϵ/σ, and polymers on the order of $N = 100$ beads. For these typical values we obtain $P_t \approx$ 1–10; this Peclet range is indicated by a shaded area in the plots of the results section (Figs. 1, 2, and 4). In contrast to this, we find that $P_t \approx 50$ would be required to model experimental dynamics. As will be shown, the discrepancy grows for greater amounts of coarse-graining (i.e., σ corresponding to larger length scales). From this, we suggest that diffusion effects are too prominent in the majority of CG simulation setups for studies of the translocation of dsDNA. One exception to this is the work of Izmitli *et al.* where the ratio of the relaxation time and translocation time of λ-DNA was evaluated in modeling the system [35].

Examining the figures in the results section, $\lambda = 50$ lies within the driven translocation regime. Hence, translocation is primarily driven and, contrary to most coarse-grained setups, diffusion is a smaller effect. Further, most simulations are being performed in the transition regime where the behavior of the probability, τ, and σ_{st} are all changing. This could greatly complicate the interpretation of the simulation results and hinder the experimental relevance.

Further, as most studies employ a freely jointed chain (FJC) model, it is unclear that they are modeling at the precision of $\sigma = 10$ nm since a FJC model of dsDNA implies that σ is at least a Kuhn length (i.e., 100 nm). Taking $\sigma = 100$ nm would have the added advantage of modeling longer DNA strands for a fixed value of N. For $\sigma = 10$ nm, polymer lengths of $N = 100$ or 200 beads (as typically studied) correspond to DNA strands 1 or 2 μm in length. On the other hand, λ-DNA, which is often used in translocation experiments [26,28,36], has a length around 16 μm and is thus about an order of magnitude longer and thus one might choose $\sigma = 100$ nm, or simulate longer chains with higher N. However, this choice also implies a nanopore with a radius of 100 nm. While experiments are performed with pores as large as this, most research focuses on tighter pores, particularly so for applications such as sizing and sequencing DNA. Further, $\sigma = 100$ nm implies an unphysically high effective DNA width of 100 nm and a large pore diameter of ≈100 nm which would not always lead to single-file translocation.

For this reason, $\sigma = 100$ nm introduces complications to the modeling that are avoided by choosing $\sigma = 10$ nm. However, since it is instructive to examine the λ required to match results for the coarser modeling, the $\sigma = 100$ nm results are also shown in Fig. 5. For this case, the λ=1.0 data underestimate the translocation Péclet number to an even greater degree, and to match the numbers a Péclet control parameter of $\lambda = 250$ is needed.

B. Single stranded DNA

Translocation of ssDNA presents a slightly more complicated scenario than dsDNA. To ensure single-file translocation of ssDNA the nanopores need to be smaller than the ones used for dsDNA. Although most people turn to biological protein nanopores for this, we will focus on work using solid-state nanopores [37–43] as these should have lower interactions with the ssDNA molecule and are closer to the "hole-in-wall" geometries studied in typical simulations. To complicate matters, ssDNA will tend to self-hybridize and form hairpins, which fundamentally changes the nature of the translocation process [41]. This could make ssDNA a bad candidate for the generic model polymer studied in the CG simulations.

Many translocation studies of ssDNA have been restricted to very short molecules (i.e., below than 50 bases) [38,39]. However, in 2005 Fologea *et al.* achieved the translocation of a 3 kb strand through a nanopore [40]. Using a pH of 13 to prevent hairpin formation from self-hybridization, they measured a translocation time of $\tau = 120$ μs using a voltage of 120 mV.

Although factors such as pH and finite size effects complicate matters, we now develop a rough estimate for the Péclet number for this experiment. Modeling this as a freely jointed chain, we can set σ to be the Kuhn length, 7 nm (note this is much more flexible than dsDNA) [44], and thus the 3 kb strand corresponds to ≈190 beads—a value that is well within the range of CG simulations [4]. Using the Kratky-Porod (worm-like chain) model, we estimate a radius of gyration of 38 nm for this strand. From Tinland *et al.* [44], we estimate the diffusion coefficient for a 3 kb strand to be $D \approx 4\ \mu\text{m}^2/\text{s}$. Using all of this in Eq. (3), we find

$$P_{\text{exp}} \approx 2. \tag{27}$$

Hence, while the P_t in typical CG setups is too low for dsDNA, it is about right for ssDNA with a length on the order of 3 kb (in agreement with the simulations of Linna and Kaski [4]). Recall that this is a relatively long strand of ssDNA for translocation studies. In the studies performed with shorter lengths [38,39], P_t will be even lower—indicating that diffusion is even more prominent. This result indicates that while translocation of dsDNA is certainly a non-equilibrium process, it may indeed be a quasi-static process for short ssDNA strands (here quasi-static indicates that the relaxation time of the polymer is much shorter than the translocation time such that the polymer is essentially relaxed at each stage of translocation) [20]. This approximation was assumed in the early theoretical work predicting scaling laws [17,18] and has been used in many theoretical and simulation studies since. While generally regarded as an unphysical limit, the above results indicate that a quasi-static process may be achievable for short ssDNA strands.

C. Rod-like viruses

One of the authors has also used the Péclet number approach to match simulation studies to experimental results for the translocation of rigid fd viruses through nanopores [45]. With a persistence length l_{p} that is greater than three times the contour length L, the viruses are rod-like and thus L is a more convenient length-scale in Eq. (3) than the radius of gyration. In this study, λ was varied between 1 and 12 for comparison to experimentally relevant voltages. As the results were found to both qualitatively and quantitatively depend on the strength of the applied field, matching the simulations to the experimental conditions was crucial for having the simulations shed light on the experimental results. The Péclet number approach allowed for a very straightforward matching of diffusion-drift effects using measurements that were easily obtainable in both experiments and simulations.

V. CONCLUSIONS

Many simulation studies of polymer translocation have employed a coarse-grained methodology. Although DNA translocation is frequently cited as a motivating application, the CG models are not often matched to experimental conditions. In this work, we propose a method for tuning one crucial aspect, the balance between drift and diffusion, to experimental conditions via a Péclet number for translocation.

Using this definition, we have demonstrated that the drift-diffusion balance in coarse-grained simulations can dramatically alter the physics of translocation. Mapping out three regimes (low, transition, and high λ values), different—and sometimes opposing results—are obtained for fundamental translocation measurements. For both in-pore and pulling forces, the probability of translocation is observed to go from very low values at low λ to 100% translocation rates at high λ. More surprisingly, the translocation time, τ, is found to vary non-monotonically across the three regimes. Finally, the variance in the translocation time is also found to be sensitive to the drift-diffusion balance.

Simulating in a regime inappropriate for comparison to experiments will thus have several consequences. First, the translocation probability will be incorrect. As most coarse-grained setups tend to over emphasize diffusion, the probability of retraction to *cis* (i.e., a failed event) has likely been overstated. Second, if we consider simulations with a wide range of polymer lengths, it is possible that short polymers will fall in one regime while long polymers fall in another. This point has been explored in several studies [1,46–49], Here we demonstrate that the drift-diffusion balance alone can introduce such difficulties. Hence, not only does this complicate the calculation of scaling exponents, but agreement between simulation and experiment could be compromised if the span of differing regimes is not the same. The non-monotonic behavior of τ with λ will thus add a complicating factor to the determination of scaling laws; the details of regime dependence invalidate the generality implied by scaling laws. This may help to explain the persistent disagreement between simulation and experimental scaling exponents. It is the subject of ongoing work to explore how the scaling exponent α derived from $\tau \sim N^{\alpha}$ depends on the Péclet number. However, here we note that a recent study has indicated that suppression of fluctuations may increase the value of α [6].

Nevertheless, as the dynamics of translocation are shown to varying significantly as the Péclet number changes, tuning the simulations to obtain the experimental correct diffusion-drift balance may be an important step in bringing simulations and experiments into eventual agreement.

Third, we investigate monomer crowding using two different force models, viz., (i) driving in the pore (with crowding) and (ii) pulling on the end (no crowding) across a wide range of Péclet numbers. We find that both yield similar behavior (for both τ and σ_{st}/τ) for a "typical" range of P_{sim}, but for high values of P_{sim}, the case with crowding shows a significant increase in both τ and σ_{st}/τ. The different results between the two models suggests that as diffusion is suppressed, monomer crowding on the *trans* side plays a larger role. The dependence of crowding effects on the Péclet number was also demonstrated for a system consisting of a polymer translocating from a nanotube through a nanopore into empty space [50]. Again, coarse-grained simulations typically overemphasize diffusion and thus these crowding effects are diminished in standard simulation setups.

Recent theoretical work has examined the impact that crowding has on the translocation process. Dubbledam et al. find that the scaling of the translocation time with polymer length approaches a universal exponent in the long-chain limit with or without crowding; in other words, crowding has no impact for very long chains [51,52]. However, the results for shorter chains are not universal and calculations with crowding yield a lower exponent than that found in the long-chain limit. Similarly, Saito *et al.* conclude that crowding introduces weak effects for finite size polymers [52]. Hence, crowding is expected to have an impact for shorter chains as found in our work. Further, crowding effects were found to have a greater impact in calculations performed at higher driving forces [51]. Since higher driving forces correspond to higher Péclet numbers, this also agrees with our finding that crowding effects play a larger role at higher Péclet numbers where diffusion is suppressed.

We have examined several different experimental translocation scenarios and estimated the Péclet P_t and corresponding Péclet control parameter λ for each case. Most significantly, we find that for translocation of dsDNA, P_t for typical simulations is anywhere from 5–50 times lower than for experimental studies. While recent progress using a tension-propagation model has united many simulation results [49], the discrepancy with experiments remains [5,7,24,25]. This may be at least partially explained by the fact that the crowding effects on the exit (*trans* side) have not been implemented in detail in this approach. Likewise, performing CG simulations at the proper P_t such that crowding effects have the correct impact may bring simulations and experiments in closer agreement.

Hence, obtaining the correct balance of drift to diffusion is crucial for modeling translocation and can be achieved this by using a Péclet number for translocation that is easily calculated for both experimental and simulation conditions. Achieving the same Péclet number in the simulations is accomplished by implementing the correct Péclet control parameter λ for a given level of coarse-graining. While the examples herein outline the effect that P_t has on the results, there are many more translocation scenarios where the impact could be felt.

ACKNOWLEDGMENTS

Simulations were performed using the ESPRESSO package [53] on the SHARCNET computer system [54] using VMD [55] for visualization. Graphics produced with MATPLOTLIB [56]. This work was funded by NSERC and the University of Ottawa.

---

[1] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, Forced translocation of a polymer: Dynamical scaling versus molecular dynamics simulation, Phys. Rev. E **85**, 041801 (2012).

[2] V. V. Lehtola, R. P. Linna, and K. Kaski, Critical evaluation of the computational methods used in the forced polymer translocation, Phys. Rev. E **78**, 061803 (2008).

[3] V. V. Lehtola, K. Kaski, and R. P. Linna, Pore-polymer interaction reveals nonuniversality in forced polymer translocation, Phys. Rev. E **82**, 031908 (2010).

[4] R. P. Linna and K. Kaski, Event distributions of polymer translocation, Phys. Rev. E **85**, 041910 (2012).

[5] K. Luo, T. Ala-Nissila, S.-C. Ying, and R. Metzler, Driven polymer translocation through nanopores: Slow-vs.-fast dynamics, Europhys. Lett. **88**, 68006 (2009).

[6] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, Driven translocation of a polymer: Fluctuations at work, Phys. Rev. E **87**, 032147 (2013).

[7] A. Bhattacharya, W. H. Morrison, K. Luo, T. Ala-Nissila, S.-C. Ying, A. Milchev, and K. Binder, Scaling exponents of forced polymer translocation through a nanopore, Euro. Phys. J. E **29**, 423 (2009).

[8] V. V. Lehtola, R. P. Linna, and K. Kaski, Unforced polymer translocation compared to the forced case, Phys. Rev. E **81**, 031803 (2010).

[9] V. V. Lehtola, R. P. Linna, and K. Kaski, Dynamics of forced biopolymer translocation, Europhys. Lett. **85**, 58006 (2009).

[10] T. M. Squires and S. R. Quake, Microfluidics: Fluid physics at the nanoliter scale, Rev. Mod. Phys. **77**, 977 (2005).

[11] T. Saito and T. Sakaue, Process time distribution of driven polymer transport, Phys. Rev. E **85**, 061803 (2012).

[12] G. W. Slater, C. Holm, M. V. Chubynsky, H. W. de Haan, A. Dubé, K. Grass, O. A. Hickey, C. Kingsburry, D. Sean, T. N. Shendruk, and L. Zhan, Modeling the separation of macromolecules: A review of current computer simulation methods, Electrophoresis **30**, 792 (2009).

[13] J. D. Weeks, D. Chandler, and H. C. Andersen, Role of repulsive forces in determining the equilibrium structure of simple liquids, J. Chem. Phys. **54**, 5237 (1971).

[14] H. W. de Haan and G. W. Slater, Mapping the variation of the translocation α scaling exponent with nanopore width, Phys. Rev. E **81**, 051802 (2010).

[15] G. S. Grest and K. Kremer, Molecular dynamics simulation for polymers in the presence of a heat bath, Phys. Rev. A **33**, 3628(R) (1986).

[16] I. Huopaniemi, K. Luo, T. Ala-Nissila, and S.-C. Ying, Polymer translocation through a nanopore under a pulling force, Phys. Rev. E **75**, 061912 (2007).

[17] W. Sung and P. J. Park, Polymer translocation through a pore in a membrane, Phys. Rev. Lett. **77**, 783 (1996).

[18] M. Muthukumar, Polymer translocation through a hole, J. Chem. Phys. **111**, 10371 (1999).

[19] H. W. de Haan and G. W. Slater, An incremental mean first passage analysis for a quasistatic model of polymer translocation through a nanopore, J. Chem. Phys. **134**, 154905 (2011).

[20] H. W. de Haan and G. W. Slater, Using an incremental mean first passage approach to explore the viscosity dependent dynamics of the unbiased translocation of a polymer through a nanopore, J. Chem. Phys. **136**, 204902 (2012).

[21] H. W. de Haan and G. W. Slater, Translocation of "rod-coil" polymers: Probing the structure of single molecules within nanopores, Phys. Rev. Lett. **110**, 048101 (2013).

[22] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, Origins and consequences of velocity fluctuations during DNA passage through a nanopore, Biophys. J. **101**, 70 (2011).

[23] D. E. Smith, T. T. Perkins, and S. Chu, Dynamical scaling of DNA diffusion coefficients, Macromolecules **29**, 1372 (1996).

[24] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, Fast DNA translocation through a solid-state nanopore, Nano Lett. **5**, 1193 (2005).

[25] P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, and D. Branton, Probing single DNA molecule transport using fabricated nanopores, Nano Lett. **4**, 2293 (2004).

[26] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker, and C. Dekker, Salt dependence of ion transport and DNA translocation through solid-state nanopores, Nano Lett. **6**, 89 (2006).

[27] V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. N. Jetha, and A. Marziali, Noise analysis and reduction in solid-state nanopores, Nanotechnology **18**, 305505 (2007).

[28] M. Mihovilovic, N. Hagerty, and D. Stein, Statistics of DNA capture by a solid-state nanopore, Phys. Rev. Lett. **110**, 028102 (2013).

[29] A. Meller, L. Nivon, and D. Branton, Voltage-driven DNA translocations through a nanopore, Phys. Rev. Lett. **86**, 3435 (2001).

[30] D. Fologea, J. Uplinger, B. Thomas, D. S. McNabb, and J. Li, Slowing DNA translocation in a solid-state nanopore, Nano Lett. **5**, 1734 (2005).

[31] V. V. Rybenkov, A. V. Vologodskii, and N. R. Cozzarelli, The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling, Nucleic Acids Res. **25**, 1412 (1997).

[32] Y. Wang, D. R. Tree, and K. D. Dorfman, Simulation of DNA extension in nanochannels, Macromolecules **44**, 6594 (2011).

[33] L. Dai, D. R. Tree, J. R. C. van der Maarel, K. D. Dorfman, and P. S. Doyle, Revisiting blob theory for DNA diffusivity in slitlike confinement, Phys. Rev. Lett. **110**, 168105 (2013).

[34] R. Adhikari and A. Bhattacharya, Driven translocation of a semi-flexible chain through a nanopore: A brownian dynamics simulation study in two dimensions, J. Chem. Phys. **138**, 204909 (2013).

[35] A. Izmitli, D. C. Schwartz, M. D. Graham, and J. J. de Pablo, The effect of hydrodynamic interactions on the dynamics of DNA translocation through pores, J. Chem. Phys. **128**, 085102 (2008).

[36] A. J. Storm, J. H. Chen, H. W. Zandbergen, and C. Dekker, Translocation of double-strand DNA through a silicon oxide nanopore, Phys. Rev. E **71**, 051903 (2005).

[37] K. Briggs, H. Kwok, and V. Tabard-Cossa, Automated fabrication of 2-nm solid-state nanopores for nucleic acid analysis, Small **10**, 2077 (2014).

[38] J. B. Heng, C. Ho, T. Kim, R. Timp, A. Aksimentiev, Y. V. Grinkova, S. Sligar, K. Schulten, and G. Timp, Sizing DNA using a nanometer-diameter pore, Biophys. J. **87**, 2905 (2004).

[39] K. Venta, G. Shemer, M. Puster, J. A. Rodríguez-Manzo, A. Balan, J. K. Rosenstein, K. Shepard, and M. Drndić, Differentiation of short, single-stranded DNA homopolymers in solid-state nanopores, ACS Nano **7**, 4629 (2013).

[40] D. Fologea, M. Gershow, B. Ledden, D. S. McNabb, J. A. Golovchenko, and J. Li, Detecting single stranded DNA with a solid state nanopore, Nano Lett. **5**, 1905 (2005).

[41] S. W. Kowalczyk, M. W. Tuijtel, S. P. Donkers, and C. Dekker, Unraveling single-stranded DNA in a solid-state nanopore, Nano Lett. **10**, 1414 (2010).

[42] M. Fyta, S. Melchionna, and S. Succi, Translocation of biomolecules through solid-state nanopores: Theory meets experiments, J. Polym. Sci. Pol. Phys. **49**, 985 (2011).

[43] M. Wanunu, J. Sutin, B. McNally, A. Chow, and A. Meller, DNA translocation governed by interactions with solid-state nanopores, Biophys. J. **95**, 4716 (2008).

[44] B. Tinland, A. Pluen, J. Sturm, and G. Weill, Persistence length of single-stranded DNA, Macromolecules **30**, 5763 (1997).

[45] A. McMullen, H. W. de Haan, J. X. Tang, and D. Stein, Stiff filamentous virus translocations through solid-state nanopores, Nat. Commun. **5**, 4171 (2014).

[46] T. Sakaue, Sucking genes into pores: Insight into driven translocation, Phys. Rev. E **81**, 041808 (2010).

[47] P. Rowghanian and A. Y. Grosberg, Force-driven polymer translocation through a nanopore: An old problem revisited, J. Phys. Chem. B **115**, 14127 (2011).

[48] T. Saito and T. Sakaue, Erratum to: Dynamical diagram and scaling in polymer driven translocation, Eur. Phys. J. E **35**, 125 (2012).

[49] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung, Unifying model of driven polymer translocation, Phys. Rev. E **85**, 051803 (2012).

[50] D. Sean, H. W. de Haan, and G. W. Slater, Translocation of a polymer through a nanopore starting from a confining nanotube, Electrophoresis, doi:10.1002/elps.201400418.

[51] J. L. A. Dubbeldam, V. G. Rostiashvili, and T. A. Vilgis, Driven translocation of a polymer: Role of pore friction and crowding, J. Chem. Phys. **141**, 124112 (2014).

[52] T. Saito and T. Sakaue, Cis-trans dynamical asymmetry in driven polymer translocation, Phys. Rev. E **88**, 042606 (2013).

[53] H.-J. Limbach, A. Arnold, B. A. Mann, and C. Holm, ESPResSo – an extensible simulation package for research on soft matter systems, Comput. Phys. Commun. **174**, 704 (2006).

[54] SHARCNET computer system, http://www.sharcnet.ca

[55] W. Humphrey, A. Dalke, and K. Schulten, Vmd - visual molecular dynamics, J. Mol. Graphics **14**, 33 (1996).

[56] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

3

Translocation of a polymer through a nanopore starting from a confining nanotube

David Sean, Hendrick W. de Haan, Gary W. Slater

I took charge of this paper: to the best of my knowledge I came up with the original idea (it was historically centred around modelling a gel's reptation tube), I designed the simulations, analysis and the theoretical adaptation. Hendrick played the role of an advisor as well as significantly contributed to the final shape of the manuscript. I wrote approximately 80% of the text.

Published in *Electrophoresis*
David Sean, Hendrick W. de Haan, Gary W. Slater: Translocation of a polymer through a nanopore starting from a confining nanotube. Electrophoresis. 2015, **36**. 682–691.

http://doi.org/10.1002/elps.201400418

Erratum

In the review process of this thesis some minor mistakes were found by the examiners. The following chapter contains a verbatim reprint of a published paper. I include the corrections in the erratum herein.

- Article page 683 second column, end of third paragraph.
  ...tension and in motion) create a ~~taught~~ taut segment that extends...

- Article page 688 first column, end of second paragraph.
  This effect contrasts with the crowding interactions...

David Sean[1]
Hendrick W. de Haan[2]
Gary W. Slater[2]

[1]Department of Physics, University of Ottawa, Ottawa, Ontario, Canada
[2]Faculty of Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Received August 29, 2014
Revised October 29, 2014
Accepted November 11, 2014

Research Article

Translocation of a polymer through a nanopore starting from a confining nanotube

In this manuscript, Langevin Dynamics simulations and Tension-Propagation theory are used to investigate the forced translocation of a polymer from a confining tube through a nanopore situated at one of the tube's ends. The diameter of the tube allows for a control over the polymer conformations: decreasing the tube diameter reduces the number of conformations available to the polymer chain both before and during translocation. As the tube diameter is decreased, the translocation time is observed to increase. Interestingly, while the width of the distribution of translocation times is reduced if the chain starts in a tube, it reaches a maximum for weakly confining tubes. A Tension-Propagation approach is developed for the tube-nanopore setup in the strongly driven limit. Good agreement between the simulations and the theory allows for an exploration of the underlying physical mechanisms, including the calculation of an effective pore friction and the assessing of the impact of monomer crowding on the *trans* side.

Keywords:
DNA in confinement / DNA sequencing / DNA sizing / Nanopore / Polymer translocation

DOI 10.1002/elps.201400418

1 Introduction

Being both ubiquitous in biological systems [1] and at the center of emerging nanotechnology such as devices for sequencing DNA, the translocation of a polymer across a membrane through a nanopore has been the subject of intense study in recent years [2–8]. For such applications, it may be desirable to control – or at least limit – the range of conformations accessible to the polymer at the start of translocation. For instance, it has been suggested that the technology-limiting wide distribution of translocation times generally observed for DNA is largely due to the initial distribution of DNA conformations [9, 10]. In this work, we explore a setup in which the polymer is initialized in a tube and subsequently translocates through a nanopore at one end of the tube. Since narrower tubes constrain the possible initial conformations to a greater degree than wide tubes, the tube diameter functions as a control parameter to explore the effect of the initial states on the translocation process.

We do indeed find significantly narrower distributions of translocation times for strongly confining tubes. Further, the mean translocation time increases as the tube diameter is reduced and thus the relative width of the distributions is much lower for narrower tubes. Interestingly, the relative width of the distributions is a maximum for weakly confining tubes before reaching an asymptotic value as the tube diameter approaches the half-space limit.

To explore these results, both Tension-Propagation (TP) theory [11–15] and Langevin Dynamics (LD) simulations [16] are used to model the dynamics. We use the strongly driven limit of TP theory – i.e. when the effects of the external field dominate over thermal fluctuations. This balance between the driven and diffusive aspects can be described by a translocation Péclet Number P_t [10, 17, 18] that we define as the ratio of the free polymer relaxation time to the translocation time: $P_t = \tau_{relax}/\tau_{trans}$. We thus limit our study to the regime where the translocation time is much shorter than the polymer relaxation time ($P_t \gg 1$), a regime that corresponds to most experimental investigations that use dsDNA [18–20]. With the free polymer relaxation time being much longer than the translocation time, the monomers follow a deterministic path set almost entirely by the initial polymer conformation [10].

In the $P_t \gg 1$ limit, good agreement is found between TP results and LD simulations. Analysis of the incremental mean first passage time (IMFPT) curves [21, 22] then allows for the calculation of an additional effective pore friction [15, 23]. To explore the effects of crowding on the *trans* side, further LD simulations are performed in which translocated monomers are removed. In the strongly driven limit, the crowding effect is shown to dominate over the pore

---

Correspondence: Professor Gary W. Slater, Department of Physics, University of Ottawa, Ottawa, ON K1N 6N5, Canada
E-mail: gary.slater@uOttawa.ca
Fax: +1-613-562-5190

Abbreviations: FENE, finitely extensible nonlinear elastic; **IMFPT**, incremental mean first passage time; **LD**, Langevin Dynamics; **TP**, Tension-Propagation; **WCA**, Weeks-Chandler-Andersen

Colour Online: See the article online to view Figs. 2, 4, 5 and 7 in colour.

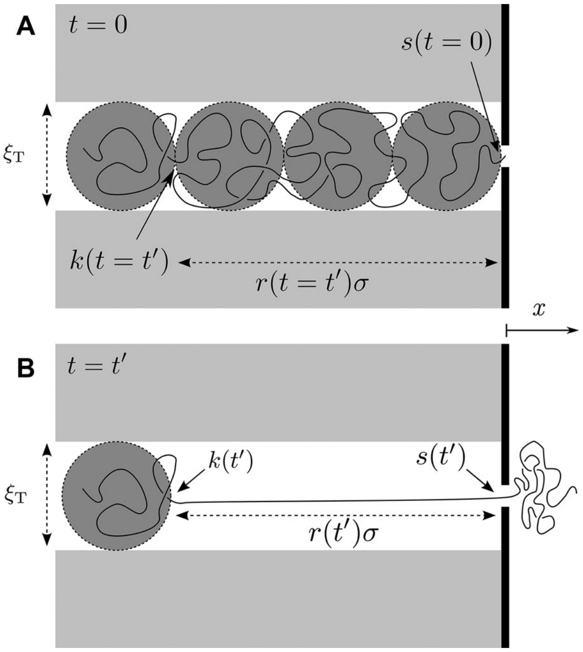


Figure 1. The polymer is confined in a cylindrical tube of diameter ξ_T on the *cis*-side. (A) The initial ($t = 0$) state where $s = k = r = 0$; overlaid are values of $k(t = t')$ and $r(t = t')$ corresponding to a later time. (B) At this later time $t = t'$, the tension-front has traveled a distance of $r(t')\sigma$ to monomer $k(t')$ while $s(t')$ monomers have translocated to the *trans*-side.

friction effects. However, we demonstrate that the balance of the two contributions shift as P_t is lowered. The manuscript is structured as follows. In the first section, TP theory is introduced and developed for the tube-nanopore geometry. Initially, pore friction and trans crowding are not explicitly taken into account in the calculations. This is followed by an overview of LD simulations that innately include both these extra contributions. Results from these two approaches are then compared to isolate the pore friction and *trans* crowding contributions.

2 Materials and methods

2.1 Tension-Propagation model in a tube

Consider a polymer initialized as an array of numbered monomers (labeled from 0 to N) of size σ with the first monomer inside the pore and the remaining monomers on the *cis*-side constrained to the accessible volume inside the tube. The semiinfinite tube is positioned such that it is concentric with the nanopore and has one end in direct contact with the membrane as shown in Fig. 1. This state can be described by the translocation coordinate s, which we take to be the index of the monomer located in the pore. Translocation begins as soon as the driving force is applied to monomer $i = 0$ inside the pore (at time $t = 0$). Under the influence of the driving force, the first monomer moves toward the *trans*-side dragging along its connected neighbors. This monomer eventually exits the pore region and stops feeling the driving force that is only applied inside the pore volume; at this point, the monomer in the pore is $i = 1$ and the translocation coordinate becomes $s = 1$.

Chain connectivity ensures that a growing number of monomers located in the tube are perturbed from their initial state as the driving force pulls in monomers through the nanopore. This *tension* propagates toward monomers down the chain as shown in Fig. 1.

The tension front $k(t)$ is defined as the monomer index at the interface (along the polymer contour on the *cis*-side) between monomers moving under the influence of the force and the remaining undisturbed monomers. In terms of the monomer indices $k(t)$ and $s(t)$, the number of *cis* monomers set in motion by the applied force is given by:

$$r(t) = k(t) - s(t). \tag{1}$$

For the situation studied here, we will consider the highly driven regime where the monomer density for the section under tension is simply one bead per unit length σ. In this description, the Pincus blob size is on the order of the monomer size. The $r(t)$ monomers being pulled on the *cis*-side (thus under tension and in motion) create a taught segment that extends down a distance of $r(t)\sigma$ away from the pore at $x = 0$.

In this highly driven regime, the distance between the tension front $k(t)$, and the pore can be estimated from the initial polymer conformation. The initial distance between monomer i and the first monomer in the pore can be used to estimate the position of the tension front once $k(t)$ reaches monomer i, as shown in Fig. 1.

We now derive a force-balance equation in the overdamped regime. Neglecting for the time being both pore friction and the (crowding) effect of translocated monomers on the *trans*-side, the effective force (we will explicitly consider frictional effects later) on a monomer in the pore, $f(t)$, is only opposed by the additional tension arising from the combined drag of the $r(t)$ monomers it needs to pull. Assuming a monomer friction coefficient of ζ, the drag force on every monomer, $-v(t)\zeta$, can be summed (in the framework of the Rouse model) to find the total drag. Furthermore, in the highly stretched regime, all the monomers under tension (including the monomer in the pore) move together at the same velocity $v(t) = \sigma ds(t)/dt$. Thus the drag force from the $r(t)$ moving monomers oppose the effective force in the pore and the force balance equation reads:

$$r(t)\sigma\zeta \frac{ds(t)}{dt} = f(t). \tag{2}$$

This equation tells us that the translocation rate $ds(t)/dt$ is limited by the ratio of the applied force to the total number of monomers under tension.

The introduction of a confining tube on the *cis*-side of the membrane affects the initial conformations of the polymer. As shown in Fig. 1, the polymer conformation can be seen as a linear sequence of de Gennes blobs the size of the tube diameter ξ_T. In the strongly driven limit where $P_t \gg 1$, this initial state can be used to estimate the length of the section under tension, $r(t = t')\sigma$, when the front reaches monomer k at time t'. By definition, all monomers beyond the tension front are unaware of the translocation action. The instant the tension front reaches a given monomer, the length of the tension line can be obtained by this monomer's x-position at $t = 0$. We can use blob theory to find the average x-position of a given monomer i, considering the initial conformation as a linear string of blobs of size ξ_{T} as shown in Fig. 1. With g monomers in each blob, there are i/g blobs between the first and i^{th} monomer. Given a blob size of ξ_{T}, the total initial distance between the monomer at the pore and monomer i can be written as $(i/g)\xi_{\mathrm{T}}$. Since inside each blob is a self-avoiding chain, we can write $\xi_{\mathrm{T}} = Ag^{\nu}\sigma$, where A is a constant of order unity for a flexible chain. We can eliminate g and write the length under tension (where i now becomes $k(t)$) as:

$$r(t)\sigma = k(t)\left(A^{1/\nu}\sigma^{1/\nu}\xi_{\mathrm{T}}^{1-1/\nu}\right). \tag{3}$$

Equations (1) and (2), and (3) can be solved to obtain the translocation rate:

$$\frac{ds(t)}{dt} = \begin{cases} \frac{f}{\zeta\sigma s}\left[A^{-\frac{1}{\nu}}(\xi_{\mathrm{T}}/\sigma)^{\frac{1}{\nu}-1} - 1\right] \text{if } s \leq s^* \\ \frac{f}{\zeta\sigma(N-s)} \qquad\qquad \text{if } s > s^* \end{cases} \tag{4}$$

where s^* is the translocation coordinate which separates the (tension) propagation and postpropagation stages. In the propagation stage, the monomer index at the front, $k(t)$, grows with time. With a finite chain, however, the tension front will eventually reach the N^{th} (and last) monomer. This occurs when the translocation coordinate s reaches the value:

$$s^* = \left[1 - A^{1/\nu}(\xi_{\mathrm{T}}/\sigma)^{1-1/\nu}\right] N. \tag{5}$$

By integrating Eq. (4) piecewise for the complete translocation process $s \in [0, N]$ the total translocation time τ is found to be:

$$\tau = N^2 \frac{\zeta}{2f}\left[A^{\frac{1}{\nu}}(\xi_{\mathrm{T}}/\sigma)^{1-\frac{1}{\nu}}\right]. \tag{6}$$

It is interesting to note that for the high-field case of tube confinement described here, the translocation time scales like $\tau \sim N^2$. Although the $\tau \sim N^2$ scaling found here was predicted by early theoretical work on translocation [24, 25], here it has a different origin and arises quite naturally from the geometry. In the quasi 1D conformation enforced by the tube, the distance the last monomer needs to travel in order to translocate scales as $\sim N$. Further, the net drag force scales with the number of dragged monomers and thus also increases as N. This yields a net scaling of $\tau \sim N^2$.

Using Eq. (4) and taking $A = 1$ for simplicity, the translocation rate given by Eq. (4) is plotted for different tube diameters in Fig. 2A. The translocation rate diverges when there are no monomers under tension, i.e. first at $s = 0$ (the start of the propagation stage) and afterwards at $s = N$ (the end of translocation process); the two hyperbolae are joined at $s = s^*$.

As the tension propagates, more and more monomers are being dragged. This is reflected as a rapid decrease in translocation rate that persists until the tension propagation reaches the last monomer ($k = N$) at $s = s^*$. Beyond s^* however, all the curves (different tube diameters) collapse to a universal one: the *cis* chain segment is entirely under tension (the tube diameter thus becomes irrelevant) and decreases in size as monomers are translocated one by one. The translocation accelerates (because the drag force on the *cis* chain is proportional to its length) until the end where the last monomer reaches the *trans*-side. Figure 2A and B also show that stronger confinement lowers the value of the critical monomer index s^* and makes the initial decrease in the translocation rate more pronounced. These effects are both due to the fact that narrower tubes extend the initial polymer conformation further from the nanopore. Close inspection of Fig. 2B reveals that the *tension-propagation* time τ_{TP} (the time difference between translocation coordinates 0 and s^*) is larger than the *postpropagation* time τ_{PP} (the time elapsed between s^* and N) for $\xi_{\mathrm{T}} = 2$. In the absence of a tube, the opposite is true [10]. This striking feature is a result of the tube geometry studied here. When the tube size is decreased, the tension propagates faster to the last monomer (decreasing τ_{TP}), which consequently means that more monomers need to be dragged for the remaining of the process (increasing τ_{PP}). The inset of Fig. 2B shows τ_{TP} and τ_{PP} as a function of ξ_{T} for different values of N. By setting the two times equal to one another ($\tau_{\mathrm{TP}} = \tau_{\mathrm{PP}}$), the analytical solution to the critical tube size $\xi_{\mathrm{T}}^* = \left(2A^{1/\nu}\right)^{\frac{\nu}{1-\nu}}\sigma$ (beyond which $\tau_{\mathrm{PP}} < \tau_{\mathrm{TP}}$), confirms that it is independent of N.

Many factors can give rise to an effective *pore friction* coefficient including an increase in viscous drag due to hydrodynamic confinement effects, an opposing electroosmotic-flow, monomer collisions with the pore edges (membrane), and attractive interactions between the pore and the monomers. The coarse-grained computer simulations utilized here will mainly introduce additional forces arising from *trans*-crowding and steric interactions (collisions) with the pore [26]. Following the work of Ikonen et al. [15, 23], we make the assumption of an additional Stokesian-like drag force $-\frac{ds}{dt}\eta_{\mathrm{P}}$, where η_{P} is some *effective* pore friction coefficient. However, we note that the contribution from collisions with the pore edges must vanish when the tube and pore diameters are identical. We thus leave the value of η_{P} as a free parameter that captures the role of collisions with the nanopore (with some dependence upon ξ_{T}) and some unknown contributions from *trans*-crowding. With this additional term, the total net force acting on the monomer in the pore can be written as:

$$f = f_{\mathrm{E}} - \frac{ds}{dt}\eta_{\mathrm{P}}. \tag{7}$$

where f_{E} is now the external force applied in the pore. Here, the net force f is not constant but depends on the rate of translocation $\frac{ds}{dt}$ (thus implicitly on the translocation

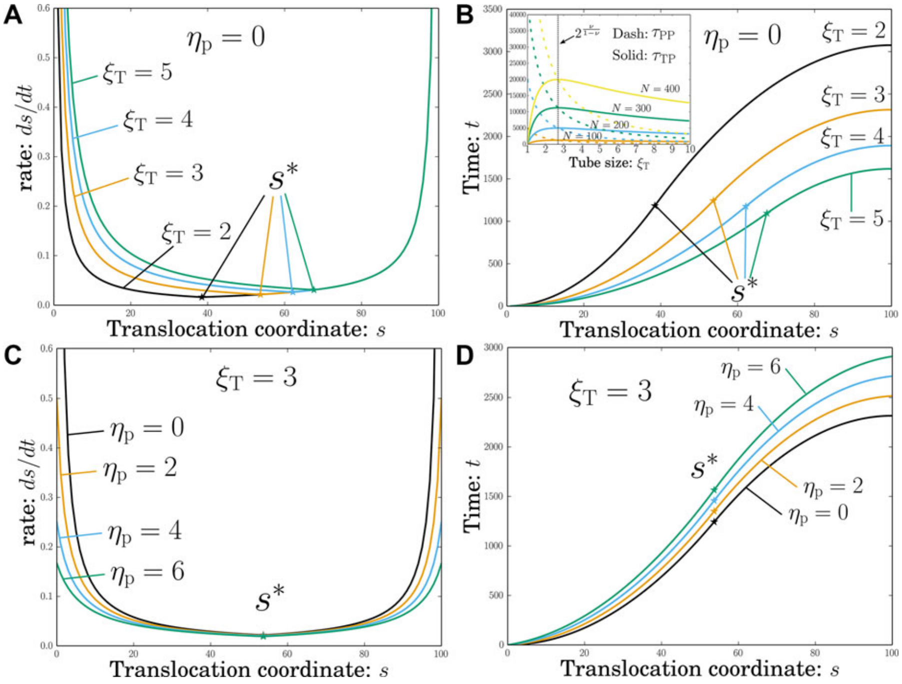


Figure 2. Results from the theoretical calculations. (A) The translocation rate ds/dt versus the translocation index s for different tube diameters ξ_T when there is no pore friction ($\eta_p = 0$). (B) The time needed to reach translocation coordinate s as a function of s for the same situation (the inset shows the tension-propagation τ_{TP} and the postpropagation time τ_{PP} as a function of ξ_T). (C) The translocation rate ds/dt versus the translocation index s for different pore friction coefficients η_p when the tube diameter is fixed at $\xi_T = 3$. (D) The time needed to reach translocation coordinate s as a function of s for the same situation. For these calculations, $N = 100$, $A = 1$ $\zeta = 1$, $\sigma = 1$, and $f_E = 1$.

coordinate s itself). Equation 4 can still be solved by separation of variables to give the total translocation time:

$$\tau = \frac{\left(A^{1/\nu}\,(\xi_T/\sigma)\,\zeta\,N + 2\,(\xi_T/\sigma)^{1/\nu}\,\eta_p\right)\,N}{2\,(\xi_T/\sigma)^{1/\nu}\,f_E} \tag{8}$$

that reduces to Eq. (6) in the low friction limit. We note that the translocation time takes a form that is in agreement with recent scaling predictions by Ikonen et al. ($\tau \sim AN^{\nu+1} + B\eta_p N$) [27] if we use $\nu = 1$, which is appropriate for our quasi-1D geometry. This result predicts the existence of a linear regime when the second term is larger, i.e. when friction is dominated by the pore and not by the chain inside the tube. For a value of $\eta_p = 3$ (consistent with the fitting value that we will be reporting later), the second term dominates only for short polymers $N < 10$, a range of polymer sizes that we do not explore here. To show the effect of this effective pore friction, Fig. 2C and D show the translocation rates and times (respectively) with different values of η_p for the case of a fixed tube diameter $\xi_T = 3\sigma$. While s^* is unaffected, translocation is obviously slowed down by this new source of friction.

2.2 Simulations

A simple coarse-grained model is used to simulate the driven polymer translocation process described in the previous section. The polymer is modeled as a generic freely jointed chain of beads linked by springs. Excluded volume interactions between the monomer beads are modeled using a Lennard–Jones potential that is truncated at the well minimum to remove the attractive Lennard–Jones component; this is also called the Weeks-Chandler-Andersen (WCA) potential and has the form [16, 28]:

$$U_{WCA}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r < r_c \\ 0 & \text{for } r \geq r_c, \end{cases} \tag{9}$$

where the well-depth ϵ is used as a fundamental energy unit and the nominal bead diameter σ as a unit for distances. The distance between two beads is r while $r_c = 2^{1/6}\sigma$ as the cutoff distance for this potential. The beads are linked to form a linear polymer by the use of finitely extensible nonlinear elastic (FENE) springs that have the form:

$$U_{FENE}(r) = -\frac{1}{2}\,k_{FENE}\,r_0^2\ln\left[1 - \frac{r^2}{r_0^2}\right], \tag{10}$$

where k_{FENE} is the spring constant and r_0 is the maximum extension. To prevent bond crossing, we use the Kremer–Grest parameters of $k_{FENE} = 30\epsilon/\sigma^2$ and $r_0 = 1.5\sigma$ [16, 29]. The constraining tube on the *cis*-side is constructed by defining a mathematical cylindrical surface with the WCA potential (Eq. 9) between the surface and the beads. Similarly, the pore is constructed by defining a planar mathematical surface (the membrane) with a hole of diameter 3σ that provides a good balance between the assurance of single-file translocation and pore friction [30].

Considering the WCA potential between these surfaces and the monomers, the space accessible to the center of each monomer is reduced. Figure 3 shows a 2D schematic of the simulation setup that depicts this reduction in accessible space. In this paper, we shall use the accessible space as the nominal metric for the size of constrictions. Therefore, point-beads have a nominal size of σ, the pore that has a mathematical diameter of 3σ will be referred to as a pore with a diameter of $d_p = 2\sigma$, a tube with a mathematical radius of 6σ will be referred as a tube of effective diameter $\xi_T = 5\sigma$,

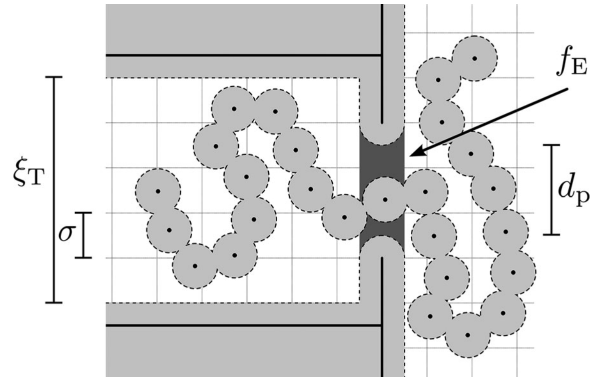


Figure 3. A two-dimensional cartoon of the simulation setup. The features are drawn to scale with the grid lines having a spacing of σ. The mathematical constraints for the pore and the confining tube are shown as solid black lines. The space inaccessible to the bead centers are shaded in light gray (with nominal edges shown with a dashed line). An external driving force is applied to beads that have their centers inside the pore (the zone shaded in dark grey). In this depiction $\xi_{\mathrm{T}} = 5\sigma$ and $d_{\mathrm{p}} = 2\sigma$.

and the thin membrane has an effective thickness (thus pore length) of σ.

The effective membrane thickness defines three key zones; the *cis*-side is defined as the volume to the left of the membrane, followed by the pore (of length σ) where the driving field is applied (dark gray shading in Fig. 3), and finally the *trans*-side to the right of the membrane.

The first monomer is positioned on the boundary between the pore volume and the *trans*-side, followed by the second that sits on the opposite boundary between the *cis*-side and the pore volume. To reduce the probability of knot formation, the remaining monomers are initialized as a straight line on the *cis*-side (thus creating a polymer of $N = 100$ monomers). The trajectory of a monomer is created by integrating the equation of motion:

$$m\vec{a} = \vec{F}^{\mathrm{C}} - \zeta\vec{v} + \vec{F}^{\mathrm{B}}, \tag{11}$$

where $\vec{a}$ is the monomer acceleration and $\vec{F}^{\mathrm{C}}$ is the sum of all conservative forces (including $\vec{f}_{\mathrm{E}}$ applied in the pore). The Brownian force $\vec{F}_{\mathrm{B}}$ has a mean of zero and a variance of $\frac{2\zeta k_{\mathrm{B}}T}{\Delta t}$, where Δt is the integration time step.

The polymer is warmed-up while keeping the first two monomers fixed to the zone-boundaries (they are allowed to move along the boundary, but cannot leave it). When the warm-up is complete, the first two monomers are released and the polymer is allowed to translocate.

Statistics on the IMFPT are accumulated by taking a timestamp whenever a monomer exits the *cis*-side for the *first* time. If the monomer ever reenters the *cis*-region by thermal fluctuations, the time is not logged when it enters the pore again. By construction, the value for the IMFPT will be 0 for the first monomer ($s = 0$) that is initialized outside of the *cis* region. The translocation event is ended when the last monomer exits the *cis* region, thus the IMFPT for the last monomer $i = 99$ is identically the mean translocation time τ. Statistics are generated by repeating this process and taking the average for 2000 independent events with different initial conformations. If all of the monomers move back to the *cis*-side, the accumulated results are discarded and the event is restarted.

We use the mean end-to-end distance R of a relaxed confined polymer at $k_{\mathrm{B}}T = 0.02\epsilon$ (data not shown) to find the prefactor $A = 1.195 \pm 0.003$ that corresponds to our simulation model. These data are also used to test the criterion for the strongly stretched regime, given as $f_{\mathrm{E}}/k_{\mathrm{B}}T > R$ [13]. We find that this is satisfied for all cases save the smallest tube (at $\xi_{\mathrm{T}} = 3\sigma$). However for this borderline case, we do not expect the results to be severely affected as the size of "flower" end of a stem-slower conformation is limited by the highly confining tube. To serve as a reference point, a polymer in free-solution (unconfined) was also simulated to obtain a mean radius of gyration of $R_{\mathrm{g}} \approx 7.38\sigma$. Unless specified, we use $k_{\mathrm{B}}T = 0.02\epsilon$ to obtain a translocation Péclet number that is relevant for experimental work done with dsDNA. We will probe how changing the Péclet number affects the results by changing the thermal activity (changing $k_{\mathrm{B}}T$) in the simulations. We thus choose a time unit independent of $k_{\mathrm{B}}T$ that we derived via the terminal velocity of a simulation bead. This yields a time unit of $[\zeta\sigma^2/\epsilon]$. In all of the simulations we use the same external driving force of $f_{\mathrm{E}} = 1[\epsilon/\sigma]$. In an effort to match the theory with the simulations, we will use the same units in both cases such that $\zeta = 1$, $\sigma = 1$, $m = 1$ and $\epsilon = 1$. In the results that follow, all variables are thus represented using units derived from ζ, σ, and ϵ and are omitted from the text for simplicity.

3 Results and discussion

3.1 Correspondence with TP theory

The resulting IMFPT for the complete translocation is plotted in Fig. 4A for different tube diameters ξ_{T}. The simulation curves share the same sigmoidal shape as the theoretical curves in Fig. 2. Using the pore friction as the only fitting parameter, the theoretical curves are shown with the simulation data. Figure 4A demonstrates that the simulated translocation dynamics are well represented by TP theory. As expected, the value of the additional friction coefficient η_{p} increases with the tube diameter; for ξ_{T} =3, 4, 5, and 6 the best fits give $\eta_{\mathrm{p}} = 0.05 \pm 0.09$, $\eta_{\mathrm{p}} = 1.56 \pm 0.09$, $\eta_{\mathrm{p}} = 1.91 \pm 0.08$, and $\eta_{\mathrm{p}} = 2.53 \pm 0.07$, respectively. The scaling predictions are verified with the log–log plot shown in Fig. 4B. Here, the translocation of various polymer lengths N is simulated for a given tube size $\xi_{\mathrm{T}} = 4$. We find a scaling of $\sim s^2$ ($\sim N^2$) on the backbone (endpoints), as well as evidence of linear scaling for low values of s.

As the tube diameter increases, the blob model assumed in the construction of the theory model will eventually break down (the polymer will no longer be represented by a linear array of blobs). We expect a transition region when the free solution radius of gyration R_{g} approaches the tube radius

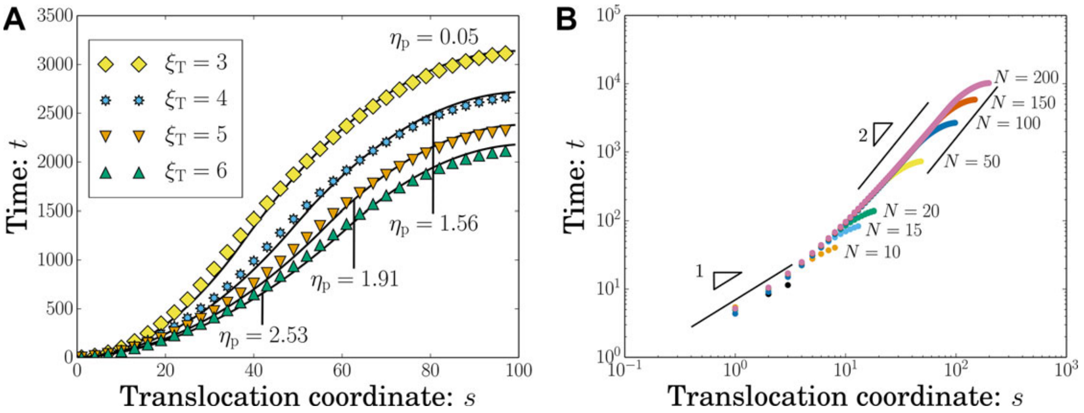


Figure 4. (A) The incremental mean first passage time to reach translocation coordinate s is plotted for different tube diameters ξ_T. The symbols represent simulation results using $N = 100$, $f_E = 1$ and $k_BT = 0.02$. The theoretical curves shown as solid lines use the stated η_p as the sole fitting parameter. (B) Log-log plot of the incremental mean first passage time to reach translocation coordinate s using various polymer lengths of N. A line with a slope of 2 is added to guide the eye next to the backbone and end-points. At low values of s the data is consistent with the predicted linear scaling along the backbone. The data shows a case with a tube size of $\xi_T = 4$, a driving field of $f_E = 1$ and a temperature of $k_BT = 0.02$.

$\xi_T \approx 2R_g = 14.7$. Near this point, these finite size effects limit the usefulness of our theoretical predictions; therefore, we do not consider these situations here, but this will be investigated in the following subsection.

We will now make two modifications to the simulations in order to explore the physical mechanisms behind the pore friction, $\eta_p v$. In these simulations, the tube size is held constant at a value of $\xi_T = 3$ (the $\xi_T = 4$ case will be explored later) while the pore size remains fixed at $d_p = 2$. Likewise, the simulation temperature is set to zero ($k_BT = 0$) such that differences between theory and simulations due to thermal fluctuations are removed (temperature effects will be investigated later).

First, the effects of crowding on the trans side are removed by deleting any monomers that are further than one monomer ahead of the monomer that is currently in the pore (for simulation stability, we always keep one monomer seeded the trans region). That is, we are deleting the trans monomers and thus artificially removing the crowding aspect. This modification, similar to what was recently done in the work of Suhonen et al. [31], is referred to hereafter as "no crowding."

Second, the effects of collisions with the pore are removed by removing the constricting pore. This is achieved by simply setting the pore diameter to match that of the tube (in this case, $d_p = \xi_T = 3$ as opposed to the usual situation where $d_p = 2$). Although the pore still exists, the constricting aspect is removed—and with it the collisions that contribute to the pore friction. For brevity, this is referred to as the "no pore" case. These two schemes yield the four simulation permutations that are shown in Fig. 5.

Since TP Theory does not explicitly include effects from either collisions with the membrane or trans crowding, the results of the "no pore, no *trans*" simulations are expected to agree with the theoretical predictions. Indeed, the theory matches the simulations without the need for an additional term (the best fit gives the small value $\eta_p = 0.23 \pm 0.04$).

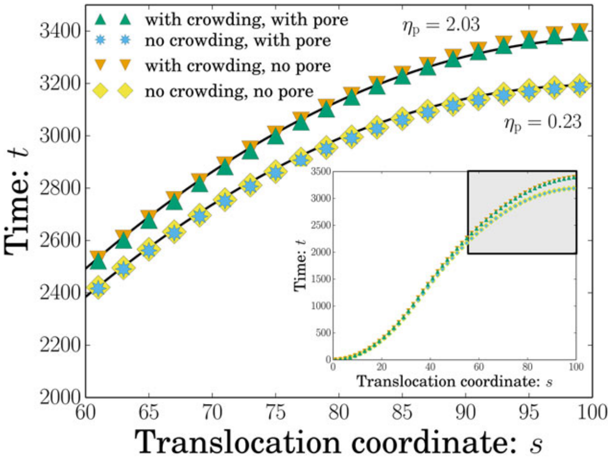


Figure 5. The incremental mean first passage time to reach translocation coordinate s is plotted versus s. With $\xi_T = 3$, different simulation schemes are used to probe different contributions to η_p. With the theoretical limit set using $k_BT = 0$, we test simulations by removing the pore membrane (to remove pore-collisions) and *trans* monomers (to remove crowding). Inset shows the time profile for the complete translocation process.

When the constricting pore is added (by using the usual diameter of $d_p = 2$), there is only a marginal effect. Visual inspection of Fig. 5 shows that the two curves essentially overlap.

We then instead look at the contribution solely from crowding (without the pore membrane). Here, the results of Fig. 5 clearly show a substantial increase in translocation time. The crowding effect can be compensated for in the theory by fitting the data using an effective pore friction that results in a value of $\eta_p = 2.03 \pm 0.03$.

Finally, the last scheme shown in Fig. 5 is for the case where both *trans* monomers and the pore membrane can play a role. Unsurprisingly, this situation yields the highest friction coefficient of $\eta_p = 2.10 \pm 0.02$. This curve is

Table 1. Changes in the translocation time arising from removing *trans* crowding, removing the narrow pore, or removing both

Tube diameter	$k_B T$	✓ pore ✓ crowding	✓ pore X crowding	X pore ✓ crowding	X pore X crowding
$\xi_T = 3$	$k_B T = 0$	τ=3315.7	–6.96 %	–0.96 %	–7.09 %
	$k_B T = 0.02$	τ=3315.1	–3.46 %	–0.39 %	–4.66 %
	$k_B T = 0.2$	τ=2221.2	–4.03 %	–2.69 %	–6.11 %
$\xi_T = 4$	$k_B T = 0$	τ=2817.0	–8.08 %	–0.78 %	–8.15 %
	$k_B T = 0.02$	τ=2657.2	–4.36 %	–0.98 %	–5.52 %
	$k_B T = 0.2$	τ=1968.8	–4.63 %	–4.98 %	–9.05 %

The percent difference is relative to the case where both the pore (with a diameter $d_p = 2$) and crowding are present (the translocation time τ given in the Table is for this case).

essentially indistinguishable from the former case due to the small impact of the membrane. Although these four simulation schemes manage to paint a convincing picture of the crowding aspect, the chosen tube size of $\xi_T = 3$ does not permit much in terms of collisions with the membrane. This is in part due to the relatively small difference between the tube diameter ($\xi_T = 3$) and the pore diameter ($d_p = 2$), and also to a subtle consequence of using $k_B T = 0$.

When $k_B T = 0$, the monomers do not stray from their initial placement until the tension has propagated down to their positions. Once under motion the monomers are aimed directly toward the center of the pore. Thus, a thermalized polymer with a center-of-mass aligned with the pore-axis will have monomers that exhibit auto-aligning motion: monomers coming from sufficiently far away from the pore will progressively approach the pore axis and be well centered once it is reached. Since there are no thermal kicks to disrupt this alignment, membrane collisions only occur during the unraveling of the first few blobs and will (almost) cease to exist for the remaining blobs. This effect contrasts the crowding interactions that increase as monomers move to the *trans*-side. Thus, among these two actors responsible for an *effective* friction—membrane collisions and *trans*-crowding—the latter has a higher relative importance near the end of the process.

To explore this physical picture, additional simulations were performed at a higher temperature of $k_B T = 0.2$. This value is chosen as it gives a balance between drift and diffusion that is typical of most coarse-grained simulations of translocations [18]. The mean translocation times τ resulting for different temperatures are shown in Table 1 for the case in which both the pore and *trans* crowding are present. The percent change in τ (with respect to pore and *trans* case) are given for the following cases: (i) with pore, no crowding (ii) with crowding, no pore and (iii) no pore, no crowding. Note that for the $k_B T = 0.2$ simulations, three monomers are seeded on the *trans* side instead of one such that the probability of successful translocation is sufficiently high (as a consequence, the observed reduction in time for this "no crowding" case will be underestimated as there will still be a small amount of crowding from these three monomers.)

As discussed previously, for $k_B T = 0$, removal of *trans* crowding significantly decreases τ, by about $\approx 7\%$. At the higher temperature of $k_B T = 0.2$, removal of *trans* crowding has a smaller effect and τ is reduced by $\approx 4\%$. Conversely, the effect of removing the pore increases with increasing $k_B T$: only a marginal decrease of $\approx 1\%$ is seen at $k_B T = 0$ while a decrease of $\approx 2.7\%$ is seen at $k_B T = 0.2$.

Hence, the impact that pore friction and *trans* crowding have on translocation depends on the Péclet number at which the system is simulated. Since a nonzero temperature allows for the crowding on *trans* to be dispersed through thermal motion, the *trans* crowding contribution decreases with increasing $k_B T$. Our results suggests that for the higher temperatures used in most published simulation studies ($k_B T \approx 1$), the *trans*-side crowding will have marginal effect, which is in agreement with the similar "no-trans" simulations of Suhonen et al. [31]

On the other hand, collisions with the pore are highly suppressed at $k_B T = 0$ (as discussed above), but have a significant effect at $k_B T = 0.2$. Thus, pore friction has an increasingly large impact as $k_B T$ increases. Finally, note that the decrease in τ when both the pore and crowding are removed is not precisely the sum of the two effects. That is, there is interplay between the effects and the net result is not additive.

When comparing across all three temperatures, the results actually appear to slightly nonmonotonic for affect of crowding, and significantly nonmonotonic for the combined effect. As demonstrated in a recent manuscript, changing the Péclet number can result in rich, nonmonotonic behavior for translocation [18]. One can hypothesize that the absolute time of translocation can play a role in the total amount of *trans* crowding. It is also worth noting that the criterion for the strongly stretching regime is not met for the higher temperatures studied here. Exploring these details is the subject of future work.

The same set of simulations was also performed for $\xi_T = 4$ to test the effect of the tube diameter. As can be seen in Table 1, the results are consistent with the discussion above: crowding effects diminish with increasing $k_B T$ while pore friction effects increase. The most notable change is that comparing between $\xi_T = 3$ and $\xi_T = 4$ for $k_B T = 0.2$, the reduction in τ on removing the pore increases from $\approx -2.7\%$ to $\approx -5\%$. For a wider tube, there is a larger cost for monomers to thread into the pore and thus pore friction

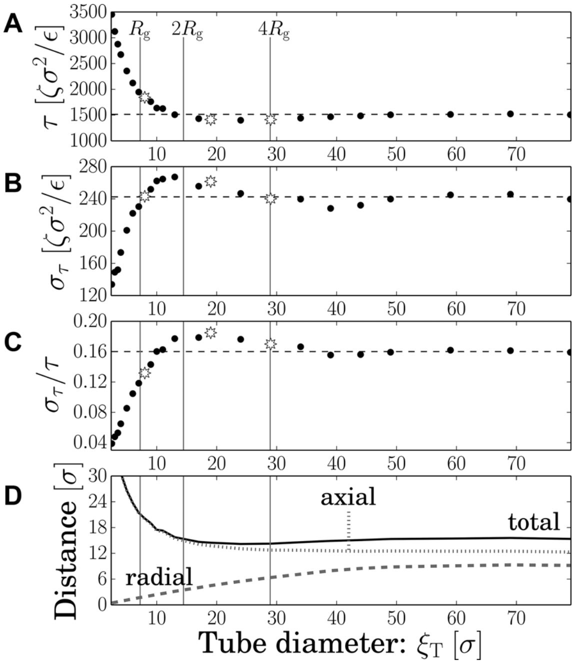


Figure 6. Four statistics of interest as a function of tube diameter ξ_T. The vertical lines show length scales of interest: R_g, $2R_g$ and $4R_g$. In (A), (B), and (C) the dashed line shows the asymptotic values for the no tube situation and the star symbols indicate the cases chosen for the monomer profiles of Fig. 7. (A) Increasing the tube size decreases the mean translocation time τ. (B) The standard deviation σ_τ as a function of ξ_T. (C) The standard deviation σ_τ rescaled by τ gives the coefficient of variation. (D) The initial monomer root-mean-squared distance from the pore. The different behavior of the radial (dashed line) and axial (dotted line) components yields a nonmonotonic curve for the total (solid line) distance.

effects are enhanced. Hence, not only do these effects depend on the value of $k_B T$, but also on the tube diameter — particularly so for the pore friction.

3.2 Half-Space transition

We now return to a Péclet number that corresponds to experimental translocation of dsDNA by restoring $k_B T = 0.02$, and we explore our main motivation, which is to probe how the tube scenario affects the translocation time distributions. The mean translocation time τ and its standard deviation σ_τ are plotted *vs* the tube diameter in Fig. 6. The values for the mean free solution radius of gyration $R_g \approx 7.4$, $2R_g$, and $4R_g$, are added in Fig. 6 as a guide. This allows us to estimate the range over which the TP theory should apply, as we expect the half-space (no tube) transition to occur beyond $\xi_T \approx 2R_g$. As expected, the width of the translocation time distributions decreases substantially for strongly confining tubes because of the restricted range of initial polymer conformations. Additionally, the strong confinement effects yield longer translocation times, as predicted by the TP theory. Both effects contribute to a notable reduction in the coefficient of variation σ_τ/τ.

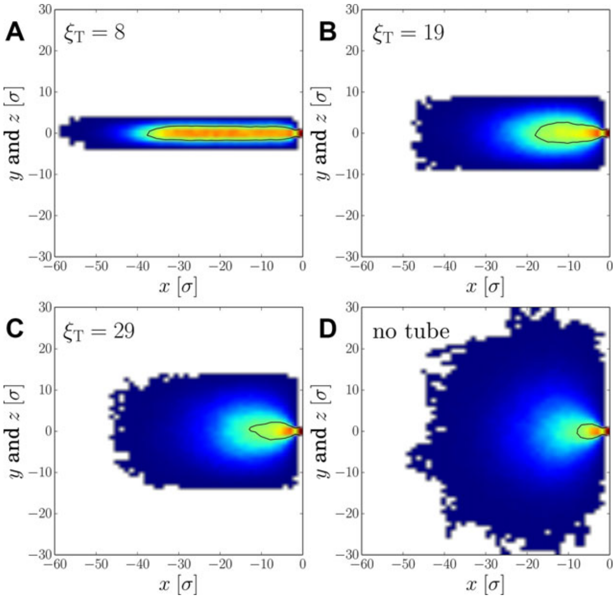


Figure 7. Two dimensional histograms of the initial monomer positions. The (x, y) and (x, z) spatial bins are overlapped into a single histogram. Four different tube diameters, ξ_T, are chosen from Fig. 6 to illustrate the confinement effects. A solid black line is added to outline a high (arbitrary) monomer count.

Remarkably, the transition between strong confinement and the half-space (no tube) is nonmonotonic. There is a minimum in τ and a maximum in σ_τ/τ before the half-space limit is reached. As we will see later, this is due to two coupled effects arising from the presence of the tube (i) *confinement* in the radial direction and (ii) *extension* in the axial direction. A better understanding of the physics can be achieved if we look at the spatial distribution of monomers inside the tube.

Figure 7 shows a 2D histogram of the initial monomer positions binned into (x, y) and (x, z). Since the pore is set at the origin, the cylindrical geometry permits us to combine z and y bins together to generate the histograms. The first panel, Fig. 7A, shows the monomer profile for a typical extended case. Here the tube is small enough to push the monomers away from the pore and translocation takes more time than the no tube case. The latter is depicted in Fig. 7D for comparison.

As the tube size is increased, monomers approach the pore while remaining somewhat confined to the inside of the cylinder. Figure 7B shows the monomer profile for a case ($\xi_T = 19$) that translocates slightly *faster* than the no-tube case. In this weak confinement case, a high number of monomers can approach the pore in the axial direction (x) while the tube keeps monomers from spreading too much in the other two directions (y and z). This creates a case where monomers are more localized near the pore, which results in a smaller mean translocation time than the half-space limit.

This hypothesis was verified plotting the root-mean-squared (where the mean $\langle\ \rangle$ is simultaneously taken across all monomers and initial conformations) axial, $\sqrt{\langle x^2\rangle}$, and radial, $\sqrt{\langle y^2\rangle + \langle z^2\rangle}$, components of the initial monomer distance to the pore, as shown in Fig. 6D. Consider first a strongly confined case: the mean axial distance is high (extended conformation) and the mean radial distance is low (compressed). As the tube diameter is increased, both components monotonically approach the no-pore limit from different directions—but at different rates (the axial decrease is steeper). Thus, when the total distance from the pore is considered (i.e. the sum), a small minimum is present, as shown in Fig. 6D. The minimum found in this analysis occurs at the same ξ_T value as it does for the mean time τ. Thus the minimum in τ is a natural consequence of distributing the monomers in the tube's cylindrical geometry while having the first monomer centered in the pore.

To uncouple the effects of τ on the standard deviation, we consider the coefficient of variation σ_τ/τ, shown in Fig. 6C. We notice that it too has a nonmonotonic behavior. The initial rise can be attributed to the reduction in initial conformations—the primary motivation for this manuscript. What is less obvious is how the σ_τ/τ can rise above the no-pore limit before slowly *decreasing*.

Since the initial increase is trivial, it suffices to motivate the decrease that occurs afterwards to justify the existence of a maximum. The decrease can be explained as follows: let us consider two tube cases that share the same translocation time. The monomer profiles for $\xi_T = 19$ and $\xi_T = 29$ are down in Fig. 7B and C, respectively. As explained previously, decreasing the tube from $\xi_T = 29$ to $\xi_T = 19$ has two effects: (i) it axially extend monomers away from the pore (which increases τ) and (ii) it radially brings monomers toward the center axis (which decreases τ). At $\xi_T = 19$, these two effects cancel one another and the translocation time is the same as for $\xi_T = 29$. Having both shorter and longer events (which average out) necessarily *increases* the standard deviation. Thus even when normalized by the time (which is the same for these two cases), the introduction of a smaller tube can yield a broader distribution. This balance of effects means that while in the strong confinement regime the tube will always yield a lower variance in translocation times, a weakly confining tube may actually increase the relative width of the translocation time distributions (compared to the no tube limit). Some of these "blob reshaping" results are reminiscent of recently published work concerning the properties of semiflexible polymers in nanochannels [32–34].

4 Concluding remarks

In this manuscript, we have investigated a polymer confined inside a tube translocating through a nanopore into free space. We have demonstrated that for strongly confining tubes, the mean translocation time is much longer and the width of the distribution of translocation times is much narrower than for the no tube case. These two effects work in concert to give a relative width of distribution of translocation times that is significantly lower for strongly confining tubes than for the usual, half-space scenario. This is a very desirable feature as it means that the size of DNA strands could be determined with greater precision in a tube-nanopore setup.

To explore these results, an extension of the Tension-Propagation theory was developed for this geometry. The results were shown to be in excellent agreement with simulations performed in the high Péclet number limit (low thermal fluctuations). Comparison between simulation data and theory allowed for the calculation of an effective pore friction, η_p, which was shown to increase as the tube radius increased. The effective pore friction arises from two factors: (i) collisions with the pore edge (membrane) and (ii) crowding of monomers on the *trans* side. Simulations were devised to explore these effects independently leading to the following conclusions: (i) the contribution of *trans* crowding is reduced as the Péclet number is decreased (temperature is increased) and (ii) friction effects due to membrane collisions increase as the Péclet number is decreased (temperature is increased). The membrane collisions were also found to increase as the tube is widened (while the pore remains the same size). Hence, the balance of factors leading to an effective pore friction depend on the Péclet number at which the system is simulated as well as on the details of the system geometry.

Although the theory presented here is limited to the strong confinement regime corresponding to small tube diameters, simulations were also performed for larger tubes up to the half-space (no tube) limit. Both the translocation time and the relative standard deviation are found to be nonmonotonic functions over the entire range of tube radii. In what we have called a weakly confining regime, the translocation time is actually a minimum while the relative standard deviation is a maximum. Examining monomer distributions and comparing the competing effects of the tube in both extending the polymer axially and confining it radially explain these effects.

Hence, a tube-nanopore setup may be useful in applications where small variations in the translocation time are desired. However, a somewhat surprising caution is that while this is true for strongly confining tubes, weakly confining tubes may actually yield a slightly wider distribution than no tube at all. Conclusions drawn from this work may also be of use for modeling systems where a gel is cast on the *cis*-side of the membrane [35,36]. The tube model presented here can be generalized to model DNA reputation through the gel prior to translocation (work in progress).

The Langevin Dynamics simulations were performed with the ESPResSo package [37] on SHARCNET (www.sharcnet.ca). Simulation visualization and plotting were done using VMD and Matplotlib [38, 39]. D.S. is supported by the NSERC-CREATE training program in Quantitative Biomedicine. This work is supported in part by an NSERC grant to G.W.S.

The authors have declared no conflict of interest.

5 References

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., *Molecular Biology of the Cell*, Garland Science, New York, 4 ed. 2002.

[2] Atas, E., Singer, A., Meller, A., *Electrophoresis* 2012, *33*, 3437–3447.

[3] Thompson, J. F., Oliver, J. S., *Electrophoresis* 2012, *33*, 3429–3436.

[4] Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J. A., *Nat. Biotechnol.* 2008, *26*, 1146–1153.

[5] Maitra, R. D., Kim, J., Dunbar, W. B., *Electrophoresis* 2012, *33*, 3418–3428.

[6] Muthukumar, M., *Polymer Translocation*, CRC Press, Boca Raton, FL 2011.

[7] Panja, D., Barkema, G. T., Kolomeisky, A. B., *J. Phys-Condens. Mat.* 2013, *25*, 413101.

[8] Dekker, C., *Nat. Nanotechnol.* 2007, *2*, 209–215.

[9] Lu, B., Albertorio, F., Hoogerheide, D. P., Golovchenko, J. A., *Biophys. J.* 2011, *101*, 70–79.

[10] Saito, T., Sakaue, T., *Phys. Rev. E* 2012, *85*, 061803.

[11] Sakaue, T., *Phys. Rev. E* 2007, *76*, 021803.

[12] Sakaue, T., *Phys. Rev. E* 2010, *81*, 041808.

[13] Saito, T., Sakaue, T., *Eur. Phys. J. E* 2011, *34*, 1–8.

[14] Dubbeldam, J., Rostiashvili, V., Milchev, A., Vilgis, T., *Phys. Rev. E* 2012, *85*, 041801.

[15] Ikonen, T., Bhattacharya, A., Ala-Nissila, T., Sung, W., *Phys. Rev. E* 2012, *85*, 051803.

[16] Slater, G., Holm, C., Chubynsky, M., De Haan, H., Dubé, A., Grass, K., Hickey, O., Kingsburry, C., Sean, D., Shendruk, T., Zhan, L., *Electrophoresis* 2009, *30*, 792–818.

[17] Dubbeldam, J. L. A., Rostiashvili, V. G., Milchev, A., Vilgis, T. A., *Phys. Rev. E* 2013, *87*, 032147.

[18] de Haan, H. W., Sean, D., Slater, G. W., *Phys. Rev. B* 2015.

[19] Storm, A. J., Storm, C., Chen, J., Zandbergen, H., Joanny, J.-F., Dekker, C., *Nano Lett.* 2005, *5*, 1193–1197.

[20] Chen, P., Gu, J., Brandin, E., Kim, Y.-R., Wang, Q., Branton, D., *Nano Lett.* 2004, *4*, 2293–2298.

[21] de Haan, H. W., Slater, G. W., *J. Chem. Phys.* 2012, *136*, 204902.

[22] de Haan, H. W., Slater, G. W., *J. Chem. Phys.* 2011, *134*, 154905.

[23] Ikonen, T., Bhattacharya, A., Ala-Nissila, T., Sung, W., *J. Chem. Phys.* 2012, *137*, 085101–085101–12.

[24] Muthukumar, M., *J. Chem. Phys.* 1999, *111*, 10371–10374.

[25] Sung, W., Park, P. J., *Phys. Rev. Lett.* 1996, *77*, 783–786.

[26] Dubbeldam, J. L. A., Rostiashvili, V. G., Vilgis, T. A., *J. Chem. Phys.* 2014, *141*.

[27] Ikonen, T., Bhattacharya, A., Ala-Nissila, T., Sung, W., *EPL-Europhys. Lett.* 2013, *103*, 38001.

[28] Weeks, J. D., Chandler, D., Andersen, H. C., *J. Chem. Phys.* 1971, *54*, 5237–5247.

[29] Grest, G., Kremer, K., *Phys. Rev. A* 1986, *33*, 3628–3631.

[30] de Haan, H. W., Slater, G. W., *Phys. Rev. E* 2010, *81*, 051802.

[31] Suhonen, P. M., Kaski, K., Linna, R. P., *Phys. Rev. E* 2014, *90*, 042702.

[32] Tree, D. R., Wang, Y., Dorfman, K. D., *Phys. Rev. Lett.* 2012, *108*, 228105.

[33] Tree, D. R., Wang, Y., Dorfman, K. D., *Phys. Rev. Lett.* 2013, *110*, 208103.

[34] Dai, L., van der Maarel, J., Doyle, P. S., *Macromol* 2014, *47*, 2445–2450.

[35] Squires, A. H., Hersey, J. S., Grinstaff, M. W., Meller, A., *J. Am. Chem. Soc.* 2013, *135*, 16304–16307.

[36] Waugh, M., Carlsen, A., Sean, D., Slater, G. W., Briggs, K., Kwok, H., Tabard-Cossa, V., *Electrophoresis* 2015, *36*, doi: 201400488.

[37] Limbach, H.-J., Arnold, A., Mann, B. A., Holm, C., *Comput. Phys. Commun.* 2006, *174*, 704–727.

[38] Humphrey, W., Dalke, A., Schulten, K., *J. Mol. Graphics* 1996, *14*, 33–38.

[39] Hunter, J. D., *Comput. Sci. Eng.* 2007, *9*, 90–95.

4

Highly driven polymer translocation from a cylindrical cavity with a finite length

Revised manuscript under review in the *Journal of Chemical Physics*

David Sean, Gary W. Slater

Highly Driven Polymer Translocation from a Cylindrical Cavity with a Finite Length

David Sean and Gary W. Slater
University of Ottawa
(Dated: December 23, 2016)

We present a computer simulation study of polymer translocation in a situation where the chain is initially confined to a closed cylindrical cavity in order to reduce the impact of conformational diversity on the translocation times. In particular, we investigate how the coefficient of variation of the distribution of translocation times can be minimized by optimizing both the volume and the aspect ratio of the cavity. Interestingly, this type of confinement sometimes increases the number and impact of hairpin conformations such that the fluctuations in the translocation process do not follow a power law in time (for instance, these fluctuations can even vary non-monotonically with time). We develop a tension-propagation model for a polymer compressed into such a confining volume and find that its predictions are in good agreement with our simulation results in the experimentally relevant strongly driven limit. Both the theoretical calculations and the simulation data yield a minimum in the coefficient of variation of the distribution of translocation times for a cylindrical cavity with an aspect ratio that makes it similar to an hemisphere. This provides guidance for the design of new devices based on the preconfinement of the target polymer into cavities.

I. INTRODUCTION

Due to both its crucial role in biological processes [1] and its promising contributions to future sequencing and diagnostic technologies [2], polymer (especially DNA) translocation has been the focus of numerous theoretical, computational and experimental studies. In fact, the availability of mono-dispersed DNA samples and the biological relevance of both single stranded and double stranded DNA make this biopolymer a very popular object of investigation for physicists; for example, the dynamics of this polymer under confinement has attracted a lot of attention over the last few years [3–9]. It is therefore no surprise that DNA translocation both into [10–14] and out of confining geometries [10, 13–18] has been the subject of intense theoretical investigation. Experimental translocation studies involving complex environments include the use of a gel matrix on different sides of the wall [19–21], translocation from micro-[22] and nano-channels [23], and trapping a DNA chain into a confining cavity between translocation events [24]. In the latter case, the DNA molecule is driven into an entropic trap where it can reside until translocating out through the very same entrance pore.

In some of our previous work, we investigated the introduction of a semi-infinite tube to confine a polymer chain prior to translocation through a nanopore [18]. This work was motivated by the observation that experimentally realized translocation experiments do not occur in a quasi-static regime [25], but rather in a highly driven regime where initial polymer conformations play a crucial role in the total translocation time (and its fluctuations) [26, 27]. This is an example of what is sometimes called "molecular individualism" [28], i.e. a physical process whose dynamics depend very strongly on the initial conformation of otherwise identical macromolecules. We showed that indeed a substantial reduction in the variance of the distribution of translocation times can be obtained if the range of initial conformations is reduced by the presence of the tube. This semi-infinite type of confinement has a single parameter: the tube diameter ξ_{T}, which serves the purpose of reducing the radial fluctuations of the initial polymer conformations. We further found that such a reduced initial conformational phase space resulted in an increase of the mean translocation times due to an increase in the mean monomer distance to the pore. The combination of these two effects leads to a win-win scenario where we obtain a much reduced coefficient of variation of the distribution of translocation times [18], i.e., a more reproducible translocation process – a highly desirable feature for the design of new instruments.

In this paper, we investigate a similar setup where the cylindrical tube now has a finite length. The additional wall placed at the end of the tube opposite the nanopore serves to limit axial fluctuations. This introduces two degrees of freedom for the confining geometry: the confinement volume v_{T} and the aspect ratio $a = L_{\mathrm{T}}/\xi_{\mathrm{T}}$, where L_{T} is the tube length. The question is thus whether there is an optimal choice of these two parameters that will narrow the distribution of translocation times even more.

However, placing a cap to prohibit elongated conformations and control axial fluctuations will also compress the polymer which can shorten the time needed to complete the translocation. Moreover, placing a cap may also increase the likelihood of having hairpins and/or knots, two types of conformations that behave quite differently.

Through the use of coarse-grained Langevin Dynamics computer simulations, we test how the mean translocation time and its standard-deviation ($\langle\tau\rangle$ and σ_τ, respectively) are affected by changing the tube size and geometry. Our goal is to find the conditions which minimize the coefficient of variation $\sigma_\tau/\langle\tau\rangle$ using the tube volume v_{T} and the aspect ratio a as the two key degrees of freedom. We complement these simulation results with theoretical arguments.

II. SIMULATIONS

In this study, we use the coarse-grained bead-spring polymer model described in our previous paper. Briefly, the monomer beads are modelled as point particles interacting via the hard-core truncated Lennard-Jones potential [29]

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r < r_{\mathrm{c}} \\ 0 & \text{for } r \geq r_{\mathrm{c}}, \end{cases} \tag{1}$$

where the nominal bead diameter σ and the parameter ϵ serve as our fundamental units of distance and energy, respectively. We use a cutoff length of $r_{\mathrm{c}} = 2^{1/6}\sigma$ such that U_{WCA} is purely repulsive for all inter bead distances r. This potential, which imposes polymer self-avoidance (point particles effectively take up a volume of σ^3), is also used for the interactions between the beads and the mathematical surfaces which define the confining geometry (in which case r is the distance between the bead centre and the given surface), namely: the confining tube; the membrane with the pore; and the end-cap on the tube. The pore of radius $r_{\mathrm{p}} = 1.4\sigma$ is chosen to be concentric with the confining tube and sits at the origin of our coordinate system – see Fig. 1.

In our coarse-grained model, the effect of the electric (driving) field is modelled as a constant force applied to those monomers whose positions are inside the nanopore. With the potential $U_{\mathrm{WCA}}(r)$ applied between the monomers and the membrane, the nanopore has an effective length of σ and thus beads with an axial (or x) coordinate which lies between $-\sigma/2$ and $+\sigma/2$ have an additional driving force F_{drive} applied to them.

To create a linear polymer (with $N = 100$ beads in our simulations), an attractive bonded potential is applied between consecutive monomers. We use the Finitely-Extensible-Nonlinear-Elastic potential

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}\, k_{\mathrm{FENE}}\, {r_0}^2 \ln\left[1 - \frac{r^2}{{r_0}^2}\right], \tag{2}$$

with $r_0 = 1.5\sigma$ and $k_{\mathrm{FENE}} = 30\epsilon/\sigma^2$ chosen to prevent bond crossing [29, 30]. With these parameters, the equilibrium root-mean-squared radius of gyration of our polymer chain is $R_{go} \approx 7.73\sigma$ and its effective volume is $v_{go} = R_{go}^3 \approx 461\sigma^3$ (in this context, equilibrium means a chain whose first monomer is fixed inside the nanopore in the limit where the tube diameter ξ_{T} and the length L_{T} are infinite).

The monomer positions are integrated in time using the Langevin equation

$$m\ddot{\mathbf{r}} = \nabla\left(U_{\mathrm{WCA}}(r) + U_{\mathrm{FENE}}(r)\right) + \mathbf{F}_{\mathrm{drive}} - \zeta\dot{\mathbf{r}} + \mathbf{F}_{\mathrm{B}} \tag{3}$$

where the terminal velocity arising from the monomer friction coefficient ζ is used to define the simulation time unit $\tau_{\mathrm{sim}} = \zeta\sigma/F_{\mathrm{drive}}$. The Brownian force $\mathbf{F}_{\mathrm{B}}$ is taken as a random vector having zero mean and variance $\langle F_{\mathrm{B}}^2\rangle = \frac{2\zeta k_{\mathrm{B}}T}{\Delta t}$, where the integration time step $\Delta t = 0.01\tau_{\mathrm{sim}}$.

We focus our study on the experimentally relevant highly-driven limit of polymer translocation. In this regime the effect of limiting the range of initial conformations should have a noticeable impact in reducing the variance of translocation times because Brownian motion, the other physical mechanism contributing to the variance, will be comparatively weak. To obtain this regime, the ratio of the driving force to thermal noise is tuned to the simulation value of $k_{\mathrm{B}}T = 0.02\epsilon$ while keeping the driving force at $F_{\mathrm{drive}} = 1.0\epsilon/\sigma$. We previously found that reducing the simulation noise term (instead of increasing the driving force) is the preferred method to obtain the desired Péclet number as it results in stable simulations in the over-damped limit [31].

The potentials described here were implemented using the simulation package ESPResSo [32]. During the simulations, we record the translocation coordinate $s(t)$ (i.e., the index of the bead that is located in the pore at time t) until we find that all monomers have translocated to the *trans*-side. The translocation time τ of an event is defined as the time required to reach the state $s(\tau) = N = 100$ for the first time. The mean translocation time $\langle\tau\rangle$ is the average translocation time for the ensemble of about 1000 different initial polymer conformations. Conversely, we also consider the statistical properties of $s(t)$ for a given time t, with an emphasis on the fluctuations of $s(t)$ at half the mean translocation time $t = \langle\tau\rangle/2$.

In certain situations we observed that translocation events took place over a remarkably long time. Confined channel geometries are known to increase the probability of polymer knot formation [33] which can cause the translocation to block or take an anomalously long time to complete [34]. When the polymer is initialized inside the tube as a self-avoiding random walk, we observe situations where knots in the polymer prohibits the continued passage of monomers (see supporting movie in Supplemental Materials [35]). We reduce the probability of forming such a knot by initializing the chain as a straight line in the tube, followed by a long equilibration stage where the end cap slowly compresses the polymer down to the desired tube length. This initialization scheme typically yields well-behaved translocations. However, outlier trajectories are still found. Visual inspection of some of these trajectories reveals that knots (mostly located near the free end of the chain) are indeed the cause for the abnormally long translocation times. Since this type of topological jamming is outside the scope of the present study, these outlier situations were removed from the analysis.

We provide a brief overview of Tension-Propagation theory in the highly driven regime in Appendix A. The translocation time can be predicted using the initial monomer coordinates obtained from the simulations as an input.

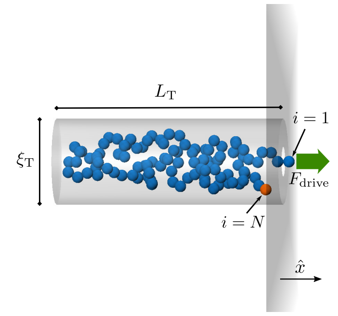


FIG. 1. (Color Online) Simulation snapshot of the initial state $s(t = 0) = 1$ for a polymer (with $N = 100$ monomers) inside a confining cylindrical cavity of length $L_{\mathrm{T}} = 19\sigma$, diameter $\xi_{\mathrm{T}} = 6.19\sigma$ and volume $v_{\mathrm{T}} = 572\sigma^3$. The nanopore has a radius $r_{\mathrm{p}} = 1.4\sigma$ and is concentric with the tube. The initial conformation shown here contains a hairpin—the last monomer ($i = N$) is spatially close to the nanopore.

III. RESULTS

A. Compression

We first investigate whether a net reduction in the variance of the translocation times, σ_τ, can be achieved by limiting the maximum extension of the chain's conformation in the axial direction. Axial compression will obviously reduce the conformational fluctuations along this direction, but it may also decrease the mean translocation time $\langle\tau\rangle$ since the monomers will on average be closer to the pore. We first test the impact of compression by simulating the system at a constant diameter ξ_{T} while reducing the tube length L_{T}.

The data in Fig. 2a shows the mean translocation time $\langle\tau\rangle$ as a function of L_{T}. Note that since the tube diameter is held constant for each curve (with $\xi_{\mathrm{T}}/\sigma = 5$, 9 or 11), decreasing L_{T} corresponds to situations of increased monomer density (or decreased tube volume v_{T}). As expected, faster translocations arise from decreasing the tube length under such conditions.

The highest confinement explored for $\xi_{\mathrm{T}} = 11\sigma$ shows a small increase of the translocation time for very short tubes ($L_{\mathrm{T}} = 2\sigma$) in Fig. 2a. Close inspection of these simulations reveal short length scale ordering of the monomers. If we only consider the axial positions of these initial conformations, we note that the monomers are pushed against the two opposite walls and thus create two monolayer stacks. This is especially apparent form the x-position histograms (data not shown) where two peaks occur directly at the wall surfaces followed by a short depletion zone before flattening out near the centre of the tube. Since *cis*-side monomer-monomer crowding restricts movement, we suspect that this effect plays in a role in the observed flattening (and increase for $\xi_{\mathrm{T}} = 11\sigma$) of the translocation time.

We also note that the standard deviations, represented as error-bars in Fig. 2a, are quite affected by the tube length. The coefficient of variation, defined as the ratio between the standard deviation σ_τ and the mean translocation time $\langle\tau\rangle$, is plotted in Fig. 2b as a function of the confining volume $v_{\mathrm{T}} = \pi(\xi_{\mathrm{T}}/2)^2 L_{\mathrm{T}}$. For the three tube diameters simulated, a minimum can be found which yields the smallest scaled standard deviation in translocation times. Again, we suspect that increasing influence of monomer-monomer collisions as the cavity volume is decreased is responsible for the minimum near $v_{\mathrm{T}} = 500\sigma^3$. As opposed to the translocation time, the coefficient of variation exhibits a clear minimum for all of the tube diameters simulated. This suggests that the standard deviation σ_τ is much more sensitive to *cis* monomer crowding compared to $\langle\tau\rangle$.

For the same three tube diameters, we also plot the Tension-Propagation predictions of Eq. A7. Using the initial monomer positions R_k from the simulations as an input, the translocation time is predicted for every independent run. The resulting coefficient of variation curves, shown in Fig. 2b are consistent with the simulation data only for volumes beyond the minimum. Of course, this theoretical model does not account for steric monomer-monomer interactions. As such, the coefficient of variation monotonically decreases with decreasing confining volume, consistent with our hypothesis that the steric interactions between monomers play a role in the observed minimum.

Since the minimum is hypothesized to arise from monomer crowding effects, the minimum should be reproducible with polymers of different length under the same critical monomer volume fraction. We tested (results not shown) the translocation of both a shorter ($N = 50$) and longer ($N = 200$) chains for the tube size of $\xi_{\mathrm{T}} = 9\sigma$ to find that indeed the minimum occurs at a similar monomer volume fraction $\phi = N\sigma^3/v_{\mathrm{T}} \approx 0.2$. Furthermore, we performed additional $N = 100$ simulations without long-range excluded volume interactions (results not shown). Simulated results from the ideal polymers follow the same monotonic behaviour as the Tension-Propagation predictions in Fig. 2b, i.e., we find no minimum when monomer crowding effects are removed.

B. Iso-Volume

In this series of simulations, the confinement volume v_{T} is held constant and the aspect ratio $a = L_{\mathrm{T}}/\xi_{\mathrm{T}}$ is varied. The scaled mean translocation time $\langle\tau\rangle/\tau^*$ is plotted as a function of a for four different cavity volumes in Fig. 3a. For these volumes, the mean translocation time decreases with decreasing v_{T} (data not shown). However, if we rescale $\langle\tau\rangle$ by $\tau^* = \tau(a = a^*)$, where $a^* = 0.436$ is the predicted position of the minimum (see next paragraph), the curves collapse near the minimum. Remarkably, the critical aspect ratio that minimizes the mean translocation time appears to be the same for all cavity volumes. It is also evident from this figure that the standard deviation σ_τ (shown as error-bars) is minimum near $a = a^*$.

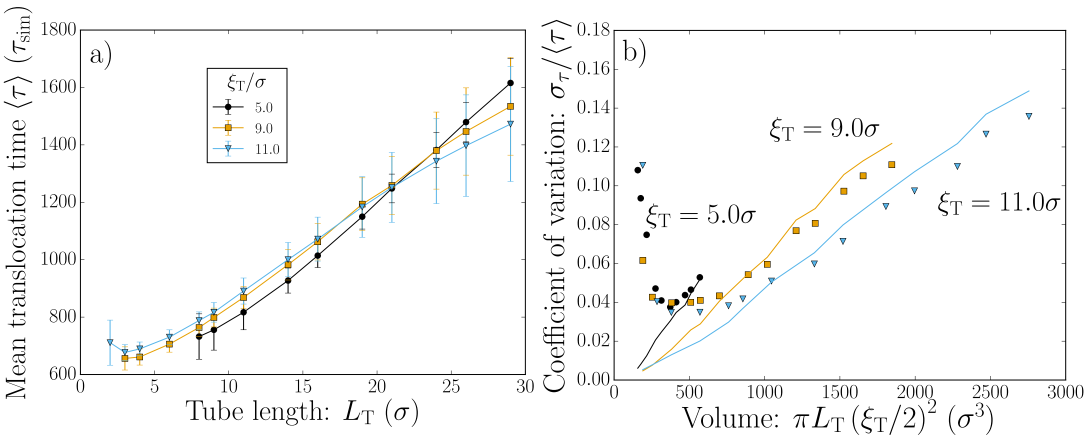


FIG. 2. (Color Online) The tube length L_{T} is varied for three different tube diameters $\xi_{\mathrm{T}} = 5\sigma, 9\sigma, 11\sigma$. a) The mean translocation time $\langle\tau\rangle$ (in units of τ_{sim}) as a function of L_{T}. The error bars indicate the standard deviation. b) The coefficient of variation $\sigma_\tau/\langle\tau\rangle$ as a function of the confinement volume v_{T}. Tension-Propagation predictions are shown as solid lines.

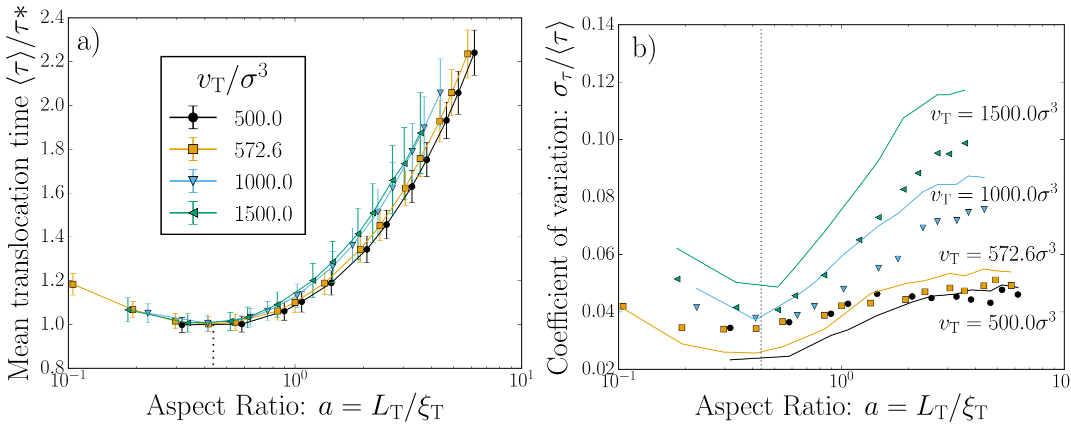


FIG. 3. (Color Online) Statistical analysis of the translocation process upon varying the aspect ratio $a = L_{\mathrm{T}}/\xi_{\mathrm{T}}$ in an ensemble where the volume v_{T} is kept constant. a) The mean translocation time $\langle\tau\rangle$ as a function of the aspect ratio. The errorbars indicate the standard deviation. The translocation time is rescaled with τ^*, which is the interpolated mean translocation time at the theoretical aspect ratio of $a^* = 0.436$, indicated by a vertical dotted line. b) The coefficient of variation $\sigma_\tau/\langle\tau\rangle$ as a function of the aspect ratio a. Tension-Propagation predictions are shown as solid lines.

For fixed-volume cavities with an aspect ratio $a > a^*$ the translocation time increases with the aspect ratio. This is because the average monomer distance to the pore increases as the geometry goes from a roughly isotropic three-dimensional cavity to a long thin tube. As was discussed previously, bringing the monomers closer to the pore will lower the translocation time. However, for an aspect ratio $a < a^*$ the translocation time also increases.

Such a reversal is not unexpected since at some point the cavity is essentially a flat disc which radially distributes the monomers further from the pore upon decreasing the axial length. With a centered pore, the geometrical crossover between the two physical pictures (long pipe vs. flat disk) is expected for an aspect ratio $a \approx 1/2$. However, this does not necessarily mean that the translocation time should be minimal precisely for this aspect ratio. Using the Tension-Propagation Theory together with a model that assumes the monomers to be evenly distributed inside the cavity, we predict that the average translocation time should be minimum for a critical aspect ratio $a^* \approx 0.436$ in Appendix B. When we rescale the data with the minimum translocation time $\tau^* = \tau(a^*)$, the curves converge to the same point at a^* for all volumes, as demonstrated in Fig. 3a.

Similarly, the standard deviation also reaches a minimum value near an aspect ratio $a = a^*$ (data not shown). In terms of the coefficient of variation, shown in Fig. 3b, the optimized geometry shares the similar critical aspect ratio a^* as the mean translocation time $\langle\tau\rangle$. We note however that the curves are very flat when the cavity volume is near the optimal value of $v_{\mathrm{T}} \approx 500\sigma^3$. The same model that predicted a minimum translocation time for

$a^* \approx 0.436$ in the previous paragraph predicts minima for $a^* \approx 0.421$ and $a^* \approx 0.414$ for the standard-deviation and the coefficient of variation, respectively.

C. Hairpins

Our study of the translocation process in the iso-volume ensemble allowed us to observe an interesting and unusual "self-focusing" phenomenon that results from the presence of hairpin conformations in the initial ensemble. In order to identify these hairpin conformations we use $x_N(t=0) = x_{oN}$, the initial axial position of the chain's N^{th} monomer as a proxy. Since the first monomer is located at $x = 0$ in the initial conformation, x_{oN} is also the x-component of the chain's end-to-end distance. The initial conformations are ranked according to their values of x_{oN} and then separated in ten groups (decile ranking). The mean x-component of the position of each of the $N = 100$ monomers in the initial polymer conformations is shown in Fig. 4a, with the corresponding histogram distribution of x_{oN} in Fig. 4b. The conformations belonging to the first decile approach the end of the tube near monomer 55, and then turn around to move towards the nanopore: these are "U"-type hairpins [9]. This behaviour is in sharp contrast to the 10^{th} decile (the polymer extends linearly towards the very end of the tube) and the mean of the full ensemble. Figure 4b shows that a substantial fraction of the polymer conformations have a "U" or "J" shape for the selected conditions as opposed to the more canonical "I"-shapes where the distance increases more-or-less monotonically with monomer index. A screenshot of a conformation identified as containing a hairpin using this metric is shown in Fig. 1. Let us now investigate how the translocation dynamics differ between the different subgroups of initial polymer conformations.

Figure 5 shows a plot of the time evolution of the mean translocation coordinate $\langle s(t)\rangle$ and its fluctuations $\langle (s(t) - \langle s(t)\rangle)^2\rangle$ as error-bars. We selected three aspect ratios and one of the fixed cavity volumes ($v_{\mathrm{T}} \approx 572.6\sigma^3$) studied in the previous section. The three sub-figures are chosen as typical cases of the: low, intermediate, and high aspect ratios (a =0.3, 3.1 and 5.8, respectively). Each time axis has been rescaled by the mean translocation time $\langle \tau(a)\rangle$ for the selected aspect ratio a. The time $t = \tau_1$ at which the first translocation of the ensemble occurs is shown in each case; note that the size of the ensemble begins to decrease after this point. In addition, we show the results from the two subgroups (the 1^{st} and 10^{th} deciles described previously) as different curves in each panel. -p

For the low and high aspect ratios (top and bottom panels, respectively), the results from the two sub-groups fall within a standard deviation of the mean. The intermediate case ($a = 3.1$, middle panel) is more striking: the fluctuations (error bars) increase much faster than in the other two cases, reach a maximum near half the process and then *decrease*; the translocation process essentially self-focuses from then on. The subgroup of conformations belonging to the 1^{st} decile ranking (hairpins) are clearly slower at first but catch up with the others near the end, whereas the (extended conformations with no hairpin) 10^{th} decile group remains well-behaved throughout. The unusual self-focusing behaviour of hairpins is consistent with tension-propagation mechanics, but is not nec-

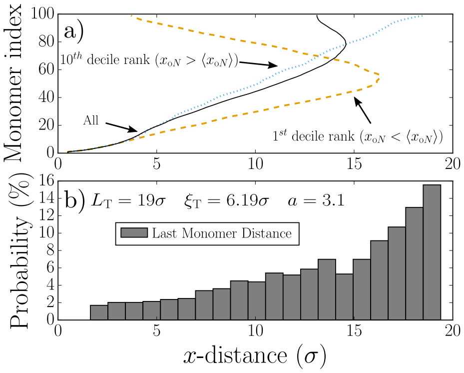


FIG. 4. (Color Online) a) The mean axial (x) distance of each of the $N = 100$ monomers is shown for the full ensemble of initial ($t = 0$) polymer conformations, as well as for the conformations in the first and last decile ranking of the initial end-to-end distances x_{oN}. b) The histogram distribution of the initial axial distances of the last monomer, x_{oN}. In these simulations, the tube has a volume $v_{\mathrm{T}} = 572.6\sigma^3$, diameter $\xi_{\mathrm{T}} = 6.19\sigma$ and length $L_{\mathrm{T}} = 19\sigma$.

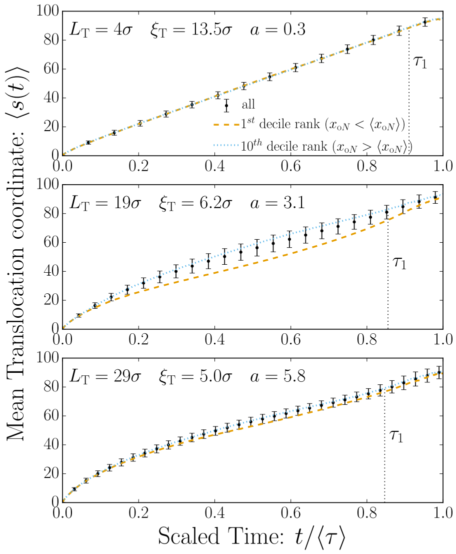


FIG. 5. (Color Online) The mean translocation coordinate $\langle s(t)\rangle$ is plotted as a function of time for three different aspect ratios a but a fixed value of the cavity volume $v_{\mathrm{T}} = 572.6\sigma^3$. The error-bars represent the standard deviation. The results from two subgroups pre-selected according to decile ranking in x_{oN} (see Fig. 4) are also shown. The time for the first complete translocation is marked as τ_1. The time axes are normalized by the mean translocation time $\langle \tau(a)\rangle$.

essarily captured by standard tension-propagation models because the latter assume that the distance between a given monomer and the pore increases monotonically with monomer index (this usually arises from using a blob model [18] and/or from pre-averaging the distances).

In the extreme case of a "U"-hairpin, the tension front propagates faster (than in the case of a "I"-conformation) at first since there are fewer "wiggles" to remove. Thus the hairpin case reaches a *slow* translocation state quite early and lags behind the blob-like (or "I"-shaped) conformations. On the other hand, this also means that the hairpins reach the accelerated stage earlier—and *before* the tension front reaches the last (i.e., the N^{th}) monomer. This permits the hairpins to rapidly "catch-up" to the blob-like conformations such that both scenarios end the process at a similar timestamp.

To fully appreciate the "self-focusing" features of $s(t)$, we plot in Fig. 6a the fluctuations $\langle (s(t) - \langle s(t) \rangle)^2 \rangle$ as a function of scaled time $t/\langle \tau \rangle$ for various aspect ratios (note that $\langle \tau \rangle$ depends upon the geometry). Since these simulations are in an iso-volume ensemble, long tube lengths (and hence large aspect ratios) correspond to situations where the tube diameter is small; it is thus important to keep in mind that hairpin conformations become increasingly rare as we increase the aspect ratio here. As can be seen in Fig. 6a, the fluctuations do not follow a simple power law and are even nonmonotonic in most cases – notably those with hairpins. The self-focusing behaviour of the fluctuations is lost when the tube is long (e.g., in the $a = 5.8$ case the maximum fluctuation is found at the time the first translocation takes place, indicated as a solid point). This is similar to the reduction of spontaneous knot formation due to restricted backfolding at high confinement [36].

At low aspect ratios where the confinement geometry resembles a flat-disc, the concept of axial hairpins become unclear and we can no longer use our decile classification in a reliable manner. In such geometries, the self-focusing feature is lost (see the $a = 0.3$ case in Fig. 6a) which suggests that hairpins are either no longer present or have become irrelevant

If we focus on the distribution of $s(t)$ at half-time $t_{\frac{1}{2}}(a) = \langle \tau(a) \rangle / 2$, which is near the time where the fluctuations reach a maximum value when self-focusing occurs (see the shaded area in Fig. 6a), Fig. 6b shows that the simulation data at the different cavity volumes nearly collapse. The mean translocation coordinate $s(t_{\frac{1}{2}}(a))$ and its standard deviation are both non-trivial functions of the aspect ratio, but are minimum when $a \approx a^*$. Interestingly, the mean monomer index varies very little over this range of aspect ratios (roughly between 50 and 57); the decrease we observe when a is reduced is again directly related to the fact that the monomers are closer to the pore when the cavity becomes roughly isotropic. The standard deviation, however, is a very sensitive function of a (it varies between about 2.7 to 6.5 here). It first increases when a is reduced because of the increase probability of having hairpins in shorter tubes. When $a \approx 2$, hairpins cannot be easily defined anymore and the standard-deviation saturates. Decreasing a further brings the monomers closer the the nanopore and the escape process becomes more deterministic,as described previously.

IV. CONCLUSION

In this paper, we have examined the physics of driven polymer translocation where the molecule starts inside a finite-volume confining tube. Our previous paper [18] showed that an open tube, which restricts the lateral extension of the polymer conformations but let the polymer chain fluctuate along the axial direction, increases the mean translocation time and substantially decreases the width of the distribution of translocation times, a favourable situation for many applications. Of course, a tube with a finite length can restrict the axial extension of the polymer conformation if it is sufficiently short. When the polymer is initialized in a compressed state (i.e., when the cavity volume is comparable to or less than the equilibrium volume of the polymer conformation), both the variance in the translocation time and mean translocation time are reduced. A compromise between these two effects was found by minimizing the coefficient of variation of the translocation time τ, which we measured for simulations of varying cavity volumes and aspect-ratios. The optimal aspect ratio was found to be $a = L_{\mathrm{T}}/\xi_{\mathrm{T}} \approx 0.5$, which corresponds to a quasi-hemispherical cavity. Although the optimal cavity volume is one which corresponds to a monomer volume fraction of ≈ 0.2, Fig. 3b shows that if one works near the optimal aspect ratio, the coefficient of variation is a rather weak function of the cavity volume. This is an important remark since it means that polymer sizes can be analysed using the same cavity as long as the aspect ratio is near optimal.

For the case of the previously studied semi-infinite tube we found that the coefficient of variation was significantly reduced by using a highly confining tube (i.e., a long and narrow tube) [18]. For instance, the lowest coefficient of variation $\sigma_{\tau}/\langle \tau \rangle \approx 0.039$ was found for the smallest tube of diameter $\xi_{\mathrm{T}} = 2.5\sigma$. In the present study, we find that a comparable reduction in the coefficient of variation can be obtained by using much wider tubes—provided that an end-cap is positioned such that the resulting aspect-ratio is near $a \approx 0.5$. As DNA loading can become experimentally challenging for highly confining tubes, these findings suggest that tuning the aspect-ratio instead of radial confinement can be used as an alternative.

For geometries that minimize the coefficient of variation, either by tuning the tube length L_{t}, or the aspect-ratio a, we find an additional feature that the mean translocation rate is rather constant throughout the process, see Appendix C. Monomer waiting times, or Incremental Mean First Passage Time data show that typical translocation dynamics exhibit a higher rate at the

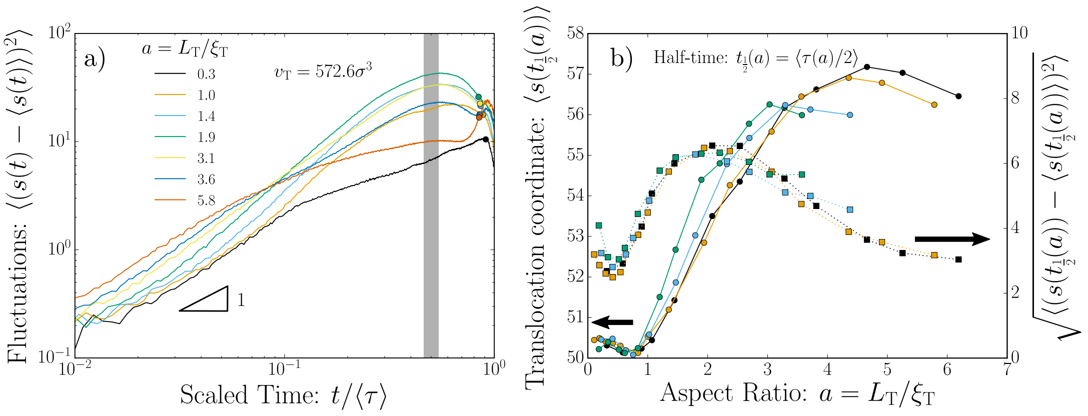


FIG. 6. (Color Online) a) Log-log plot of the fluctuations of $s(t)$ vs scaled time $t/\langle \tau(a) \rangle$. For every tube aspect ratio a shown here, the time for the first chain to translocate, $t = \tau_1$, is indicated by the circle marker. b) We now take a slice at the half-time $t_{\frac{1}{2}}(a) = \langle \tau(a) \rangle/2$ and show both the mean (circles, solid line) and the standard deviation (square, dotted line) of the translocation index s as a function of the cavity aspect ratio for four different cavity volumes.

beginning and at the end of the process. Tuning the pre-confinement geometry as a way to obtain a constant translocation rate can be beneficial in nanopore sequencing applications.

Statistics of the translocation coordinate s for a given time t were also studied. We found that placing a polymer in a finite tube can yield conformations containing hairpins; these conformational features map out into a rich temporal landscape where fluctuations can actually *decrease* with time – an interesting example of self-focusing. For these situations our simulation data do not follow a power law of the type $\langle \Delta s^2 \rangle \sim t^\beta$ [37–39], and thus we cannot define a β exponent.

For the case of a moderately compressed polymer, we showed that we can model the process by treating the polymer as a space-filling gas of monomers. Under this assumption, we calculated the mean monomer distance to the pore, $\langle R \rangle$. Using the tension-propagation theory in the highly-driven limit, together with $\langle R \rangle$, we were able to correctly predict the role of the cavity aspect ratio. This theoretical model predicts a minimum in the coefficient of variation in R for an aspect ratio ≈ 0.5, which is consistent with our simulations. These calculations, together with our simulation data, predict that a hemispherical cavity (with the pore placed at the centre of the circular base) should lead to the narrowest distribution of translocation times.

In conclusion, translocating a polymer from a hemispherical cavity whose volume results in a monomer volume fraction of ≈ 0.2 is predicted to be the optimal condition for those situations where polymer translocation is used as a way to measure the polymer length.

V. ACKNOWLEDGEMENTS

Data was plotted with Matplotlib [40] and ESPResSo [32] simulations were visualized in VMD [41]. This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Grant RGPIN/046434-2013 to GWS and the University of Ottawa.

Appendix A: Tension-Propagation Theory

The following describes the tension-propagation model and is mostly based on the approach proposed by Sakaue, and Saito and Sakaue [42–45]. We follow a construction and variable name scheme similar to our previous work [18]. Let s be the monomer index in the pore, k the monomer index at the tension front and $R_k\sigma$ the distance between monomer k and the pore. In the high force regime, the polymer segment between the tension front and the pore forms a straight line, thus the distance R_k is simply given as

$$R_k = (k - s)\sigma, \tag{A1}$$

with its time derivative

$$\frac{dR_k}{dt} = \left(\frac{dk}{dt} - \frac{ds}{dt} \right) \sigma. \tag{A2}$$

The translocation rate is determined by the balance of the drag force due to the $(k - s)$ moving monomers (we neglect monomers on the *trans*-side) and the external force applied to the monomer located in the pore:

$$\zeta \sigma \, (k - s) \, \frac{ds}{dt} = F_{\mathrm{d}}, \tag{A3}$$

where ζ is the friction coefficient of one monomer. These three equations can be solved by separation of variables to yield the time required for the tension front to reach the last monomer ($k = N$) which we call the *tension-propagation* time (subscript TP):

$$t_{\mathrm{TP}} \frac{F_{\mathrm{d}}}{\zeta} = \int_0^N R_k dk - \frac{R_N^2}{2\sigma}. \tag{A4}$$

During the second stage of the translocation process, the so-called *post-propagation* stage (subscript PP), the tension front remains at $k = N$ and Eq. A2 becomes

$$\frac{dR_k}{dt} = -\frac{ds}{dt}\sigma. \tag{A5}$$

We can separate the variables again and integrate over $s \in (s^*, N)$, where $s^* = N - R_N/\sigma$ is the value of s at the beginning of the post-propagation stage. This yields

$$t_{\mathrm{PP}} \frac{F_{\mathrm{d}}}{\zeta} = \frac{R_N^2}{2\sigma}. \tag{A6}$$

The total translocation time is thus given by

$$\tau = t_{\mathrm{TP}} + t_{\mathrm{PP}} = \frac{\zeta}{F} \int_0^N R_k dk. \tag{A7}$$

This corresponds to the the strongly-stretched limit of Eq. (11) in ref. [45]. Note that including the contribution of t_{PP} ends up cancelling the last term in Eq. A4 here. Different driving regimes yields a similar scaling form, but with a different exponent for the driving force F.

Appendix B: Homogenous monomer density approximation

Although monomer connectivity was needed in establishing the integral in Eq. A7, its evaluation (a sum for our discrete model) can be performed in an arbitrary manner if we are only interested in the final translocation time. However, this is not true for any other time of interest like the ones shown in Fig. 6 for hairpins. Thus for situations where monomer-monomer collisions and entanglement effects can be neglected, all that is needed are the initial distances between each monomer and the pore. For a free chain, this means averaging over all possible initial conformations. In the case of confined geometries where the polymer is compressed, however, the polymer actually occupies the entire confining volume with essentially a uniform monomer density. In other words, we can solve the problem by treating the initial polymer chain as a gas of free monomers.

For a given cavity geometry, one can easily calculate the expected (or mean) distance $\langle R \rangle$ between a randomly selected location in the cavity and the pore, and from this we can predict the average translocation time of a compressed polymer using the simple expression

$$\langle \tau \rangle = (\zeta/F) N \langle R \rangle. \tag{B1}$$

This approach is obviously only valid if the cavity volume v_{T} is smaller than the polymer volume R_{go}^3. Note that calculating $\langle R \rangle$ averages-out the molecular individualism.

For a cylindrical cavity as used in this paper, we must compute the integral

$$\langle R \rangle = \frac{1}{v_{\mathrm{T}}} \int_{\phi=0}^{2\pi} \int_{\rho=0}^{\xi_{\mathrm{T}}/2} \int_{z=0}^{L_{\mathrm{T}}} R \rho d\phi d\rho dz, \tag{B2}$$

where $R = \sqrt{\rho^2 + z^2}$. In the iso-volume ensemble, we can write the solution to the above integral in terms of a the confining volume v_{T} and the aspect ratio $a = L_{\mathrm{T}}/\xi_{\mathrm{T}}$. With this change of variables, finding the aspect ratio a^* which minimizes R for a fixed volume involves solving for the root of

$$\sqrt{4a^2+1}\left(32a^3 + 2a\right) - 3\mathrm{arcsinh}(2a) - 64a^4 = 0, \tag{B3}$$

which we can numerically obtain as $a^* \approx 0.436$. This is fully consistent with the simulation data presented in this paper. The spatial results arising from the homogenous monomer density model agree with the spatial results from simulated initial conformations, thus verifying that the approximation holds for the volumes studied.

The spatial variance $\sigma_R^2 = \langle R^2 \rangle - \langle R \rangle^2$ does not directly map onto the temporal variance $\langle \tau^2 \rangle - \langle \tau \rangle^2$. A scaling argument for the temporal variance is given in [45]. However, the simulation results of both σ_R^2 and σ_τ^2 (data not shown) show that both curves share the same qualitative features, namely, both have a minimum at the same position. It is interesting to note that the minima in σ_R and $\sigma_R/\langle R \rangle$ occur at aspect ratios of $a \approx 0.421$ and $a \approx 0.414$ respectively, also consistent with our (temporal) simulation results.

For a tube with a square cross-section, $\langle R \rangle$, σ_R^2 and $\sigma_R/\langle R \rangle$ are all minimum at the same aspect ratio of $a^* = 1/2$. It is easy to show that it is for the truly isotropic hemispherical cavity that the ratio $\sigma_R/\langle R \rangle$ is the smallest (0.258) when we compare to the optimally shaped (i.e., using $a = a^*$) cylindrical (0.277) and square section (0.289) tubes: unsurprisingly, the confining geometry with the potential of yielding the lowest coefficient of variation is thus the hemisphere, but the three geometries differ by only a few percent.

Appendix C: Incremental Mean First Passage Times

The dynamics of the translocation process can also investigated by plotting the Incremental Mean First Passage Time (IMFPT) as a function of the translocation coordinate [46, 47]. These plots highlight how the rate of translocation varies during the process. In typical scenarios, translocation starts at a high rate which decreases until reaching a minimum point after which it accelerates up to the end of the process; this type of behaviour is reflected in IMFPT plots as sigmoidal curves. These dynamics are well modelled with the Tension-Propagation theory (see Appendix A), where the rate is determined by the number of monomers set in motion by the driving force [18]. Initially, only the first monomer is affected by the force and the translocation rate is fairly high. As translocation progresses however, the tension propagates down the polymer and more monomers drift towards the pore under the influence of the driving force; this increases friction and decreases the translocation rate. The tension eventually reaches the last monomer which is on average the furthest from the pore. The resulting stretch of moving monomers is responsible for the slowest translocation rate. After this point the translocation accelerates until the last monomer exists the pore. This "typical" translocation picture for the iso-volume ensemble can be seen as the sigmoidal curve found for $a = 5.78$ in Fig. 7a. We choose to show only results in the iso-volume ensemble for brevity; the data with constant tube diameters are similar. Note that non-monotonic mean translocation time of Fig. 3a are showed as the last points $s = 100$ of Fig. 7a.

The effect of having a geometrical constraint, however, can strongly affect the shape of these curves. For some geometries the variation in translocation rates are significantly reduced which results in translocation dynamics which are, on average, more uniform in time. We perform a linear least-squares (χ^2) fit to the data and use the root-mean-squares of the residuals to quantify the linear *vs* sigmoidal rate profiles.

Figure 7a and b shows that the cavity aspect-ratio where coefficient of variation is minimized near $a \approx 1/2$ results in a situation where the average translocation rate is the remarkably constant throughout. In light of tension-propagation dynamics, this suggests that the tension-segment being restricted by the cavity does not significantly vary, on average, over the course of the whole translocation process. These results demonstrate that translocation from these cavities not only yield a minimum variation in mean translocation times, but also increases uniformity of the average translocation *rate* throughout the whole process. Both of these features could be used for new applications.

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2002).

[2] M. Muthukumar, *Polymer Translocation* (CRC Press, Boca Raton, 2011).

[3] A. Khorshid, P. Zimny, D. Tétreault-La Roche, G. Massarelli, T. Sakaue, and W. Reisner, Phys Rev Lett **113**, 268104 (2014).

[4] E. Werner, F. Persson, F. Westerlund, J. O. Tegenfeldt, and B. Mehlig, Phys Rev E **86**, 041802 (2012).

[5] L. Dai, J. van der Maarel, and P. S. Doyle, Macromolecules **47**, 2445 (2014).

[6] A. R. Klotz, L. Duong, M. Mamaev, H. W. de Haan, J. Z. Y. Chen, and W. W. Reisner, Macromolecules **48**, 5028 (2015).

[7] A. R. Klotz, M. Mamaev, L. Duong, H. W. de Haan, and W. W. Reisner, Macromolecules **48**, 4742 (2015).

[8] A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman, J Chem Phys **140**, 084905 (2014).

[9] P. Cifra and T. Bleha, Soft Matter **8**, 9022 (2012).

[10] M. Muthukumar, J Chem Phys **118**, 5174 (2003).

[11] K. Zhang and K. Luo, Soft Matter **9**, 2069 (2013).

[12] K. Zhang and K. Luo, J Chem Phys **140**, 094902 (2014).

[13] J. M. Polson, M. F. Hassanabad, and A. McCaffrey, J Chem Phys **138**, 024906 (2013).

[14] J. M. Polson, J Chem Phys **142**, 174903 (2015).

[15] P. J. Park and W. Sung, Phys Rev E **57**, 730 (1998).

[16] M. Muthukumar, Phys Rev Lett **86**, 3188 (2001).

[17] A. Cacciuto and E. Luijten, Phys Rev Lett **96**, 238104 (2006).

[18] D. Sean, H. W. de Haan, and G. W. Slater, Electrophoresis **36**, 682 (2015).

[19] A. H. Squires, J. S. Hersey, M. W. Grinstaff, and A. Meller, J Am Chem Soc **135**, 16304 (2013), pMID: 24143914.

[20] M. Waugh, A. Carlsen, D. Sean, G. W. Slater, K. Briggs, H. Kwok, and V. Tabard-Cossa, Electrophoresis **36**, 1759 (2015).

[21] Z. Tang, Z. Liang, B. Lu, J. Li, R. Hu, Q. Zhao, and D. Yu, Nanoscale **7**, 13207 (2015).

[22] R. Tahvildari, E. Beamish, V. Tabard-Cossa, and M. Godin, Lab Chip **15**, 1407 (2015).

[23] Y. Zhang and W. Reisner, Nanotechnology **26**, 455301 (2015).

[24] X. Liu, M. M. Skanata, and D. Stein, Nat Commun **6** (2015).

[25] Y. Kantor and M. Kardar, Phys Rev E **69**, 021806 (2004).

[26] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, Biophys J **101**, 70 (2011).

[27] J. Sarabadani, T. Ikonen, and T. Ala-Nissila, J Chem Phys **141**, 214907 (2014).

[28] P. de Gennes, Science **276**, 1999 (1997).

[29] G. W. Slater, C. Holm, M. V. Chubynsky, H. W. de Haan, A. Dubé, K. Grass, O. A. Hickey, C. Kingsburry, D. Sean, T. N. Shendruk, *et al.*, Electrophoresis **30**, 792 (2009).

[30] G. S. Grest and K. Kremer, Phys Rev A **33**, 3628 (1986).

[31] H. W. de Haan, D. Sean, and G. W. Slater, Phys Rev E **91**, 022601 (2015).

[32] H.-J. Limbach, A. Arnold, B. A. Mann, and C. Holm, Comput Phys Commun **174**, 704 (2006).

[33] C. Micheletti and E. Orlandini, Soft Matter **8**, 10959 (2012).

[34] A. Rosa, M. Di Ventra, and C. Micheletti, Phys Rev Lett **109**, 118301 (2012).

[35] Supplemental Material.

[36] C. Micheletti and E. Orlandini, ACS Macro Lett **3**, 876 (2014).

[37] J. Chuang, Y. Kantor, and M. Kardar, Phys Rev E **65**, 011802 (2001).

[38] J. Dubbeldam, A. Milchev, V. Rostiashvili, and T. A. Vilgis, Europhys Lett **79**, 18002 (2007).

[39] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, Phys Rev E **85**, 041801 (2012).

[40] J. D. Hunter, Comput Sci Eng **9**, 90 (2007).

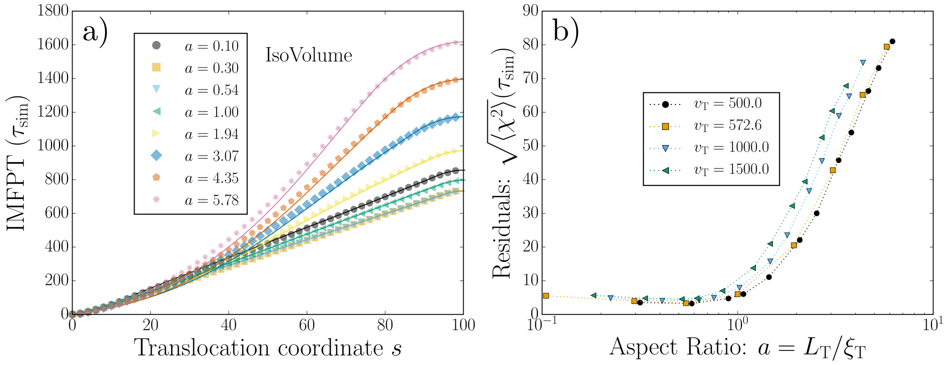


FIG. 7. (Color Online) a) The mean time needed to reach a translocation coordinate s for the first time is plotted vs the translocation coordinate s. Keeping the volume constant at $v_{\mathrm{T}} = 572.6\sigma^3$ the aspect ratio $a = L_{\mathrm{T}}/\xi_{\mathrm{T}}$ is varied. The solid lines are numerical solutions to TP-theory scaled with the last simulation point τ b) The quality of a linear fit is shown as the root-mean-squared of the residuals. For all volumes considered, the IMFPT curves are relatively linear near the aspect ratio of $a \approx 0.5$.

[41] W. Humphrey, A. Dalke, and K. Schulten, J Molec Graphics **14**, 33 (1996).
[42] T. Sakaue, Phys Rev E **76**, 021803 (2007).
[43] T. Sakaue, Phys Rev E **81**, 041808 (2010).
[44] T. Saito and T. Sakaue, Euro Phys J E **34**, 1 (2011).
[45] T. Saito and T. Sakaue, Phys Rev E **85**, 061803 (2012).
[46] H. W. de Haan and G. W. Slater, J Chem Phys **134**, 154905 (2011).
[47] H. W. de Haan and G. W. Slater, J Chem Phys **136**, 204902 (2012).

5

Reducing the variance in the translocation times by stretching the polymer

Hendrick W. de Haan, **David Sean**, Gary W. Slater

This paper was started by Hendrick when he was a Postdoctoral Fellow with Gary. I became involved after Hendrick had generated the simulations and began the analysis. I performed my own simulations and analysis—I wanted to also profile the forces against the membrane as well as save the initial conformations. Following Gary's suggestion I re-ran the simulations and varied both F_d and $k_\mathrm{B}T$ (the original simulation had only varied $k_\mathrm{B}T$). I contributed to the writing of the manuscript and provided the analysis for the time evolution of σ_τ. I wrote approximately 40% of the text. I have explicit permission from Hendrick to include this manuscript as a chapter in this thesis.

To be submitted shortly to the *Journal of Chemical Physics*

Reducing the variance in the translocation times by stretching the polymer

Hendrick W. de Haan
Faculty of Science, University of Ontario Institute of Technology,
Oshawa, Ontario, Canada, L1H 7K4

David Sean, Gary W. Slater
Physics Department, University of Ottawa,
Ottawa, Ontario, Canada, K1N 6N5

(Dated: December 25, 2016)

Abstract

Langevin Dynamics simulations of polymer translocation are performed where the polymer is stretched via two opposing forces applied on the first and last monomer before, and during, translocation. In this setup, polymer translocation is achieved by imposing a bias between the two pulling forces such that there is net drift towards the *trans*-side. Under the influence of pre-stretching forces, the elongated polymer ensemble contains fewer variations in conformations compared to an unstretched ensemble. Simulations demonstrate that this reduced spread in initial conformations yields a reduced variation in translocation times relative to the mean translocation time. This effect is explored for different ratios of the amplitude of thermal fluctuations to driving forces to control for the relative influence of the thermal path sampled by the polymer. Since the variance in translocation times is due to contributions coming from sampling both thermal noise and initial conformations, our simulations offer independent control over the two main sources of noise, and allow us to shed light on how they both contribute to translocation dynamics. Experimentally relevant conditions are highlighted and shown to correspond to a significant decrease in the variation of translocation times, thus indicating that stretching DNA prior to translocation could assist in nanopore-based sequencing and sizing applications.

PACS numbers: 87.15.ap, 82.35.Lr, 82.35.Pq

Keywords: Nanopore, translocation, polymer, simulations, Peclet number, variation

I. INTRODUCTION

The translocation of polymers across membranes through nanopores is central in biological processes at the cellular level [1] and in the development of DNA analysis techniques [2]. Applications such as nanopore-based sequencing technologies have led to numerous experimental [3–8], theoretical [9–17], and computational [18–24] studies of polymer translocation. Experimental histogram distributions of the translocation times τ exhibit surprisingly large variances — particularly so when considering the remarkable monodispersity of DNA. This large variance in the translocation time is a prominent source of uncertainty for applications such as using nanopores for DNA size determination. It also introduces complications for sequencing applications where a consistent and orderly passage of bases is desirable. Hence, techniques to reduce this variation are of great interest in the development of nanopore technologies. In this work we propose and explore a methodology for reducing this variation by prestretching the polymer prior to translocation.

In an idealized translocation setup that does not include effects such as polymer-wall interactions, the degradation of the nanopore or interactions with ions or impurities in the solvent, the variance in translocation time is known to come from two different contributions: i) Brownian noise arising from thermal fluctuations (i.e., the *stochastic path* that a polymer follows in a particular translocation event) and ii) conformational noise arising from otherwise identical polymers starting translocation in different conformations [11, 25–27]. Hence, the variance in the translocation time can be reduced by either reducing the Brownian noise (e.g., turning down the temperature) or by reducing the variation between the different initial conformations (i.e., starting all translocation events from very similar initial states).

In this work, we narrow the distribution of initial conformations by stretching the polymer. Prior to the translocation, the polymer is stretched by the application a force to the first and last monomer in opposite directions. As the stretching force is increased, conformations in which the polymer starts compressed near the pore or even in a relaxed state become less and less likely. In the limit of a very strong stretching force, all events would start with the polymer in a linear conformation along the pore axis. Reducing conformational noise by confining the polymer in a long cylinder [27] or a cylindrical cavity [28] prior to translocation has been investigated previously. However, the use of this dual pre-stretching force can reduce complications like knots and hairpins.

The stretching force, F_{s}, thus enables us to implicitly control the range of initial conformations while Brownian fluctuations can be controlled via the simulation variable $k_{\mathrm{B}}T$. Once the polymer is relaxed in a pre-stretched state, the translocation process is initiated by increasing the force applied on the end of the polymer initially located in the pore by F_{d} such that there is a net drift towards the *trans*-side through the nanopore. Driving the translocation via end-pulling eliminates the need to consider the effects of monomer crowding on the *trans*-side of the membrane, yielding a clean setup to investigate translocation dynamics.

II. THE STRETCHING-PULLING FORCE SETUP

Figure 1 shows a schematic of how both a stretching force and a driving force are implemented in our simulations. A stretching force F_{s} is applied to the first monomer ($i = 1$) towards *trans* while the same force is applied to the last monomer ($i = N$) in the opposite direction. Although the force F_{s} stretches out the polymer, the net force on the polymer remains zero. Hence, to drive the translocation process, an additional pulling force F_{d} is applied to the first monomer. This pulling force setup corresponds to translocation as controlled by optical tweezers or an attached magnetic bead [29–31]. While this has been previously studied via simulations [21], the new feature here is the combination of the stretching and pulling forces. A previous simulation study also employed a double force arrangement but in that work, a force in the pore opposed the pulling force [32].

Lehtola et al. [33] investigated how the initial polymer configuration affects the scaling law behaviors by simulating a polymer chain with an initial configuration of monomers in a straight line. They observed a scaling $\tau \sim N^{\alpha}$ with an exponent of $\alpha = 2$, which is the same as the driven limit of Sung and Park [9] where the monomer-fluid friction dominates over the monomer-nanopore friction. These findings indicate that, not only the applied force and the length of the chain, but also the initial polymer conformation affects the scaling exponent strongly.

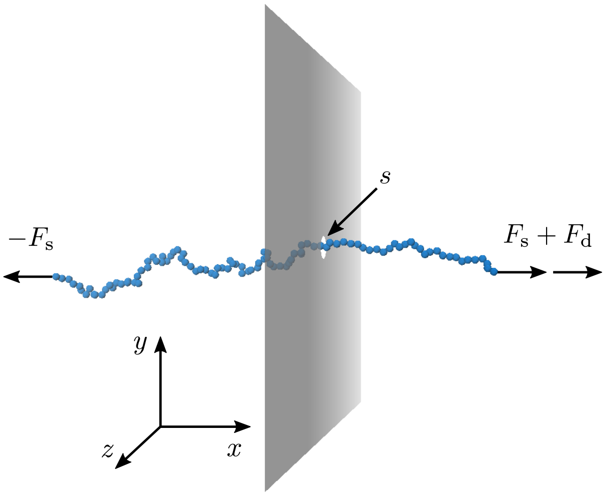


FIG. 1: (Color online) In this force configuration, the polymer is stretched by applying two equal and opposite forces $\pm F_{\mathrm{s}}$ on the polymer ends. An additional force F_{d} is applied to the first monomer to drive the polymer to the *trans*-side.

A. Polymer simulations

A setup that is very close to a standard coarse-grained Langevin Dynamics (LD) simulation approach is used to model the system [34]. The excluded volume interactions between the $N = 100$ monomer beads are implemented via the WCA potential [35]:

$$U_{\mathrm{WCA}}(r) = \begin{cases} 3\left(4\epsilon_{\mathrm{LJ}}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon\right) & \text{for } r < r_{\mathrm{c}} \\ 0 & \text{for } r \geq r_{\mathrm{c}}. \end{cases} \tag{1}$$

Here the nominal well depth ϵ_{LJ} serves as a fundamental energy scale, and the nominal bead diameter σ as the fundamental length scale. Unless otherwise noted, we report length, energy and force variables in units of σ, ϵ_{LJ}, and $\epsilon_{\mathrm{LJ}}/\sigma$ respectively. The cutoff distance is set to $r_{\mathrm{c}} = 2^{1/6}\sigma$ such that this potential is purely repulsive.

To connect monomers along the length of the polymer, the finitely extensible nonlinear elastic (FENE) potential is used. Note that for this work, as relatively high stretching forces are used, we multiply these two standard potentials (WCA and FENE) by a factor of three in order to reduce bond-stretching artifacts:

$$U_{\mathrm{FENE}}(r) = -3\left(\frac{1}{2}kr_0^2\ln\left(1 - \frac{r^2}{r_0^2}\right)\right). \tag{2}$$

To prevent bond crossing [36] the parameters are chosen to be $k = 30\epsilon_{\mathrm{LJ}}/\sigma^2$ and $r_0 = 1.5\sigma$.

Monomer positions are integrated in time using the Langevin equation of motion [34]. Monomers are subject to random thermal kicks which have a variance $2\zeta k_{\mathrm{B}}T/\Delta t$, where $\Delta t = 0.01$ is the integration time step. Since we will vary the magnitude of thermal fluctuations via the variable of $k_{\mathrm{B}}T$, we report temporal quantities in units of $\zeta\sigma^2/\epsilon_{\mathrm{LJ}}$.

The membrane is modelled as a mathematical surface with a pore of radius 1.5σ. The beads and the membrane interact via the WCA potential. The nanopore thus has an effective diameter of 2σ and a length of σ.

The external stretching force $\pm F_{\mathrm{s}}\hat{x}$ is applied on the first and last monomer (respectively) and the polymer is equilibrated. However, since the tension blob size scales as $\sim k_{\mathrm{B}}T/F_{\mathrm{s}}$, decreasing (increasing) the value of $k_{\mathrm{B}}T$ for a given stretching force F_{s} will yield an increased (decreased) end-to-end distance. To decouple this change in the amount of pre-stretching upon variations in the control parameter $k_{\mathrm{B}}T$, the polymer is stretched by pulling on both ends with a force $F_{\mathrm{s}} = \hat{F}_{\mathrm{s}}k_{\mathrm{B}}T/\sigma$. This ensures that the dimensionless stretching force $\hat{F}_{\mathrm{s}} = F_{\mathrm{s}}\sigma/k_{\mathrm{B}}T$ yields the same degree of deformation. Thus, when we report a particular value for $\hat{F}_{\mathrm{s}}$, this will always amount to the same degree of polymer deformation, irrespective of the value of $k_{\mathrm{B}}T$ used in the simulations.

The warmup procedure consists of pre-stretching the polymer with $-F_{\mathrm{s}}$ whilst having the first monomer confined in the pore. After the warmup, the monomer held in the pore is released and pulled by additional external force $F_{\mathrm{s}} + F_{\mathrm{d}}$. The simulation is re-started with a new initial conformation whenever the polymer is found to be completely on the *cis*-side of the membrane; this is considered to be a failed translocation attempt. A total of 1000 successful translocations are generated for each data point.

In a recent manuscript, we demonstrated that the dynamics of polymer translocation through a nanopore are significantly affected by the balance between the magnitude of the force driving the polymer towards the *trans*-side, and the diffusive aspects arising from thermal noise [37]. To quantify this relationship, one can define the *translocation Péclet* number as the ratio between the polymer relaxation time and its translocation time τ [11, 37, 38]:

$$P_t = \frac{1}{\tau}\frac{R_{g0}^2}{D_0} \approx \frac{F_{\mathrm{d}}}{kT}, \tag{3}$$

where R_{g0} and D_0 are the free solution radius of gyration and diffusion coefficient of the polymer respectively. In that work, we found that typical coarse-grained (CG) simulation setups, where the force and diffusion coefficient are on the order of unity in LD units, yield a P_t that is too low for modelling realistic dsDNA translocation events. Instead, a ratio of $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T \approx 50$ is appropriate for a model where $\sigma = 10$ nm such that σ matches the effective width of dsDNA and is comparable to the pore diameters used in experiments. However, typical CG models employ a freely-jointed chain where, as $\sigma = 10$ nm implies, the persistence length is $\approx 5\ \sigma$. On the other hand, if one sets $\sigma = 100$ nm such that σ corresponds to a Kuhn length thereby justifying the use of a freely jointed chain, a ratio of $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T \approx 250$ should be used.

The two relevant energy scales are thus the thermal energy $k_{\mathrm{B}}T$ and the energy associated with monomers crossing the nanopore, $F_{\mathrm{d}}\sigma$, where σ is the effective width of the membrane. In the current simulations, we map the effect of pre-stretching over a wide range of $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T$ ratios — from 1 to 100. Values of interest are thus: $F_{\mathrm{d}}\sigma/kT \approx 1-10$, which corresponds to typical CG setups; $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T \approx 50$ which corresponds to a good (albeit too flexible) model of dsDNA; and $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T \geq 100$, which corresponds to even more coarse-grained models.

In order to cover a range where $F_{\mathrm{d}} \in (1-100)k_{\mathrm{B}}T/\sigma$ while remaining both in the overdamped limit (F_{d} cannot be arbitrarily high) and being able to obtain events in a reasonable simulation time (F_{d} cannot be arbitrarily low), both F_{d} and $k_{\mathrm{B}}T$ are varied. Simulations at driving forces of $F_{\mathrm{d}}\sigma/\epsilon_{\mathrm{LJ}} = 0.1$, 0.2, 0.5, and 1.0 are performed for two values of the thermal energy: $k_{\mathrm{B}}T = 0.01\epsilon_{\mathrm{LJ}}$ and $k_{\mathrm{B}}T = 0.10\epsilon_{\mathrm{LJ}}$. To compare across cases, we define a dimensionless driving force $\hat{F}_{\mathrm{d}} = F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T$. Note that the cases overlap at $\hat{F}_{\mathrm{d}} = 10$ — attained using the two combinations: i) $F_{\mathrm{d}} = 0.1\epsilon_{\mathrm{LJ}}/\sigma$ with $k_{\mathrm{B}}T = 0.01\epsilon_{\mathrm{LJ}}$; and ii) $F_{\mathrm{d}} = 1\epsilon_{\mathrm{LJ}}/\sigma$ with $k_{\mathrm{B}}T = 0.1\epsilon_{\mathrm{LJ}}$ — which allows us to verify that there are only minimal differences between these cases and that the effects overwhelmingly arise from the value of $\hat{F}_{\mathrm{d}}$.

III. RESULTS

A. Translocation times τ

Figure 2 shows the translocation time distributions at different pre-stretching forces $\hat{F}_{\mathrm{s}} = 0, 1, 5, 25$ for three driving forces $\hat{F}_{\mathrm{d}}$. Note that for these plots as well as others in the present manuscript, the translocation time τ is scaled by the driving force F_{d}. The absolute

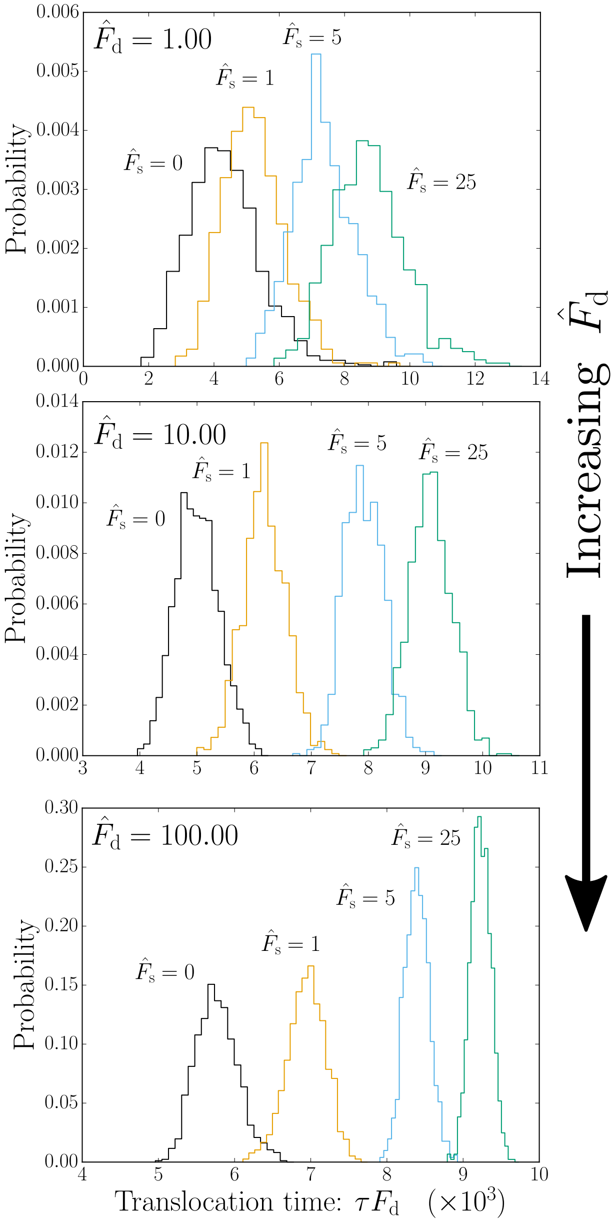


FIG. 2: (Color online) Histogram distributions of the translocation time with different pre-stretching forces $\hat{F}_\mathrm{s}$. The three plots show different driving regimes ($\hat{F}_\mathrm{d} = 1$, $\hat{F}_\mathrm{d} = 10$, and $\hat{F}_\mathrm{d} = 100$).

translocation time is trivially shorter for cases with a higher driving force; the rescaling used here assumes that $\langle \tau \rangle \sim F_\mathrm{d}^{-1}$, which is not exactly the case, as we shall see later. Examining the qualitative differences between the $\hat{F}_\mathrm{d}$= 1, 10, 100 panels, one notes that the overlap between the distributions diminishes with increasing driving magnitude $\hat{F}_\mathrm{d}$. Recall that the source of variance has contributions from both thermal noise and initial conformations;

increasing $\hat{F}_\mathrm{d}$ mostly reduces the contributions from the thermal noise, whereas increasing $\hat{F}_\mathrm{s}$ reduces the contributions from initial conformations. Thus the segregation between the $\hat{F}_\mathrm{s}$ histograms as $\hat{F}_\mathrm{d}$ is increased arises from reducing the effect of thermal noise. Conversely, the $\hat{F}_\mathrm{d}$=100 panel clearly showcases how conformational noise is suppressed with pre-stretching. A striking feature here (and at all simulated $\hat{F}_\mathrm{d}$ values), is that the mean translocation time $\langle\tau\rangle$ increases considerably with increasing $\hat{F}_\mathrm{s}$.

This is shown more clearly in Fig. 3 for all $\hat{F}_\mathrm{d}$ values. Again, the mean translocation time $\langle\tau\rangle$ is rescaled by the force F_d. The data in Fig. 3 exhibits two plateaus: at low $\hat{F}_\mathrm{s}$ with no pre-stretching and at $\hat{F}_\mathrm{s} \to \infty$ where the initial conformations approach a rod-like conformation. At $\hat{F}_\mathrm{s} = 0$, the slight increase of $\langle\tau\rangle F_\mathrm{d}$ with increasing $\hat{F}_\mathrm{d}$ highlights how the driving regime affects the translocation time. Both datasets corresponding to $\hat{F}_\mathrm{d} = 10$ (with different combinations of $k_\mathrm{B}T$ and F_d) give the same result, demonstrating that the simulations are indeed following the same the physical process.

In the high prestretching limit $\hat{F}_\mathrm{s} \to \infty$ the conformations all approach a rod-like initial conformation. We expect the data to asymptotically collapse for the different $\hat{F}_\mathrm{d}$ values here since the $\langle\tau\rangle \sim 1/F_\mathrm{d}$ should be strictly valid for rods, as confirmed by the data.

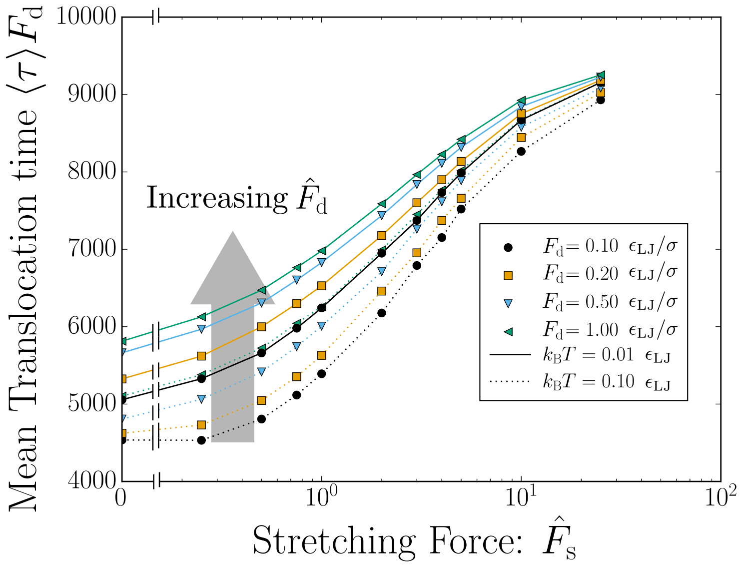


FIG. 3: (Color online) The average translocation $\langle\tau\rangle$ time scaled by the driving force F_d as a function of the scaled stretching force $\hat{F}_\mathrm{s}$.

Two effects need to be highlighted in order to understand the increase in $\langle\tau\rangle$ with $\hat{F}_\mathrm{s}$. First, we need to consider the tension-propagation (TP) translocation dynamics while taking into account the fact that tension propagates faster along a stretched polymer. As the polymer is pulled through the pore the net force is simply the pulling force F_d applied to the first monomer. However, the net drag on the polymer is proportional to the number of monomers that have been set into motion; monomers outside the range of the tension front do not contribute to the drag [11–15]. When $\hat{F}_\mathrm{s}$ is low and the polymer is not in a stretched conformation, the number of monomers being dragged increases as translocation proceeds, but it is initially quite low as the pulling force is pulling out the slack in the coil rather than dragging the entire coil. Conversely, when the stretching force is very high, the polymer starts nearly fully extended and thus more monomers contribute to the drag earlier; the tension then front quickly reaches the end of the polymer at which point all monomers contribute to the drag. This picture is consistent with the tension propagation model and has recently been explicitly tested for the case of driven polymer translocation from a tube [27, 28]. The data plateau at high $\hat{F}_\mathrm{s}$ indicates the saturation of the underlying polymer deformations.

For large $\hat{F}_\mathrm{d}$ values, the driving force dominates over thermal forces, hence the evolution of the polymer is largely deterministic: the relative unimportance of thermal fluctuations means that the polymer does not relax as translocation progresses, i.e., the process is highly out of equilibrium.

Conversely, at small $\hat{F}_\mathrm{d}$ values, the polymer is allowed to partially adapt as translocation proceeds. The driving force deforms the polymer via tension propagation which effectively moves the center of mass of the *trans*-monomers *away* from the nanopore. If the polymer relaxes somewhat from this non-equlibrium state, it does so by primarily having monomers move *towards* the pore. Hence, as the thermal energy is increased, the effective drag is reduced and the translocation time decreases. This is seen in Fig. 3 at low $\hat{F}_\mathrm{s}$ where the higher $k_\mathrm{B}T$ curves are below the lower $k_\mathrm{B}T$ curves and within each $k_\mathrm{B}T$ case, higher driving forces yield larger translocation times. As the stretching force $\hat{F}_\mathrm{s}$ increases, the impact of the non-equilibrium effects is diminished. At high $\hat{F}_\mathrm{s}$ values, the chain is nearly completely stretched and acts almost as a single body for the entire translocation process regardless of the balance between F_d and $k_\mathrm{B}T$. Consequently, very little deviation of τ between $\hat{F}_\mathrm{d}$ values is observed in this limit and the data converges. Note that the $\hat{F}_\mathrm{s} = 0$ case is studied in detail in ref [37] where we demonstrate that this additional increase in τ at low $\hat{F}_\mathrm{s}$ reflects varying degrees of non-equilibrium effects.

B. Translocation time variations σ_τ

Returning to Fig. 2, the variation in the widths σ_τ of the distributions depend not only on $\hat{F}_\mathrm{s}$ but also on $\hat{F}_\mathrm{d}$. In the last panel with $\hat{F}_\mathrm{d} = 100$, corresponding to the driving force dominating over thermal forces, the width σ_τ decreases with increasing $\hat{F}_\mathrm{s}$. For $\hat{F}_\mathrm{d} = 1$ (first panel) and $\hat{F}_\mathrm{d} = 10$ (middle panel), the behaviour is less obvious and the width is only weakly dependent upon $\hat{F}_\mathrm{s}$. Comparing across panels, σ_τ decreases with increasing $\hat{F}_\mathrm{d}$ at any particular $\hat{F}_\mathrm{s}$ value. This reflects the suppression of diffusion resulting in reduced variation of the stochastic paths and narrower distributions.

The value of σ_τ is plotted against $\hat{F}_\mathrm{s}$ for different $\hat{F}_\mathrm{d}$ values in Fig. 4a. As expected, σ_τ decreases with $\hat{F}_\mathrm{s}$ at large $\hat{F}_\mathrm{d}$ values. At $\hat{F}_\mathrm{d} = 1$, the width initially decreases but subsequently increases while at $\hat{F}_\mathrm{d} = 2$ and 5, σ_τ is essentially independent of $\hat{F}_\mathrm{s}$.

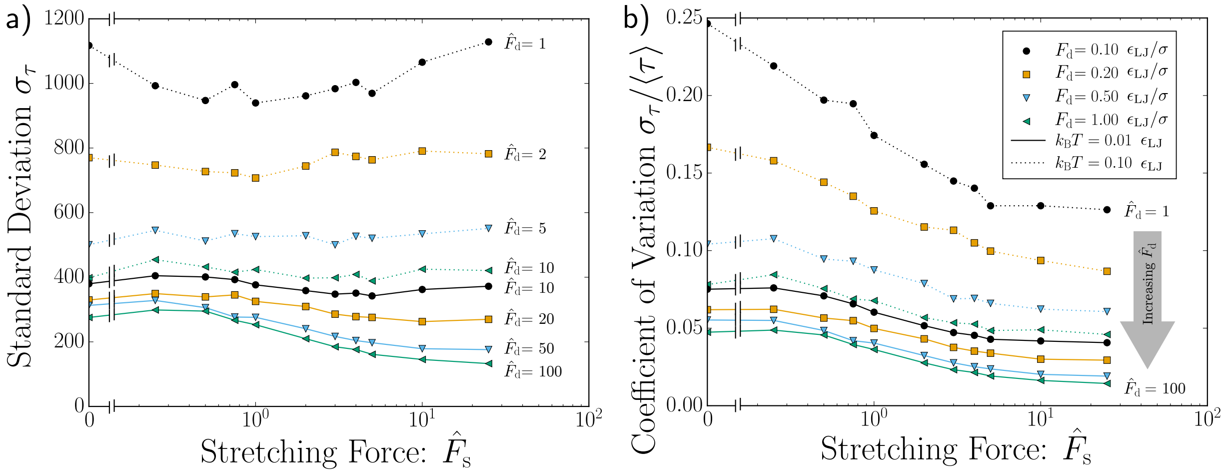


FIG. 4: (Color online) a) Standard Deviation σ_τ and b) Coefficient of variation $c_v = \sigma_\tau/\langle\tau\rangle$ plotted as a function of the stretching force $\hat{F}_\mathrm{s}$ for several values of the driving force $\hat{F}_\mathrm{d}$.

As discussed, there are two contributing factors to the width of the distribution of translocation times: the ensemble of initial conformations and the effects of polymer diffusion during the translocation process. One expects that the latter mechanism, the variation due to diffusion, to grow with time. Recalling that the translocation time increases significantly as the stretching force increases, this means that there are two competing effects for the data shown in Fig. 4a. Due to the fact that the mean translocation time $\langle\tau\rangle$ increases with $\hat{F}_\mathrm{s}$, diffusion will obviously cause σ_τ to increase with $\hat{F}_\mathrm{s}$. On the other hand, increasing $\hat{F}_\mathrm{s}$ reduces the ensemble of initial conformations and this causes a reduction in σ_τ with increasing

$\hat{F}_\mathrm{s}$.

The competition between these effects yields the behaviour seen in Fig. 4a. At low $\hat{F}_\mathrm{d}$ and low $\hat{F}_\mathrm{s}$, σ_τ is relatively independent of $\hat{F}_\mathrm{s}$ (excluding the non-monotonic case of $\hat{F}_\mathrm{d} = 1$) indicating that these two effects largely balance out. This is largely true also for the case of high $\hat{F}_\mathrm{d}$ and low $\hat{F}_\mathrm{s}$.

However, the behaviour at high $\hat{F}_\mathrm{s}$ depends on $\hat{F}_\mathrm{d}$. At low $\hat{F}_\mathrm{d}$, σ_τ increases with increasing $\hat{F}_\mathrm{s}$ indicating that the diffusive effects dominate over the further reduction in the range of initial conformations. On the other hand, at high $\hat{F}_\mathrm{d}$, σ_τ decreases with increasing $\hat{F}_\mathrm{s}$ and hence diffusive effects are marginal compared to the narrowing of the distributions that arises from the reduced conformational phase space. Low $\hat{F}_\mathrm{d}$ thus corresponds to diffusion dominated dynamics while for high $\hat{F}_\mathrm{d}$ the process is primarily driven.

C. Time evolution of σ_τ

Up to this point, we have been interested in the statistics at a specific time, namely the end of the translocation process where the translocation coordinate s equals N and the time t equals τ. Thus, the focus has been on the accumulated noise σ_τ. However, the physical picture put forth by the tension-propagation theory suggests that when an end-pulling process is considered [11], the noise arising from the initial conformations plateaus when the tension front reaches the last monomer. After this stage the N monomers of the polymer move in unison with a constant velocity, and the source of fluctuations are solely Brownian. Thus the way by which the fluctuations grow should exhibit the characteristic two-steps of the underlying dynamics. To explore this, we examine statistics of the translocation coordinate s as a function of time t.

Log-log plots of the variance $\sigma_s^2 = \langle (s - \langle s \rangle)^2 \rangle$ as a function of the scaled time $t/\langle \tau \rangle$ are shown in Fig. 5. In Fig. 5a we show the $\hat{F}_\mathrm{s} = 5$ case alone with two additional curves selected to highlight the contrast between fluctuations arising from Brownian noise *versus* those due to initial conformations. To do this, we perform a set of deterministic simulations where the polymer is initiated with conformations according to $\hat{F}_\mathrm{s} = 5$ but $k_\mathrm{B}T$ is set to zero such that there is no thermal noise. Hence, the resulting fluctuations arise solely due to variations in the initial conditions. These deterministic data are shown as a labeled dashed line in Fig. 5a. To contrast this, we also plot the expected fluctuations arising from Brownian

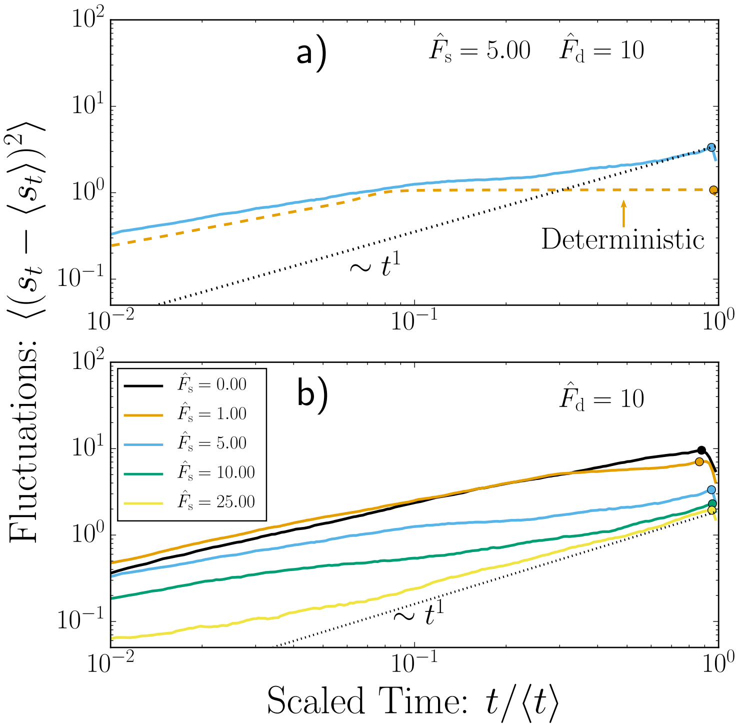


FIG. 5: (Color online) Fluctuations of s vs scaled time $t/\langle\tau\rangle$. The pre-stretching situations are shown as different curves. The circle symbol dot near $t = \langle\tau\rangle$ indicates the time for the first translocation of the ensemble, i.e., the population decreases from this point onwards.

motion, modelled as a line with slope unity ($\sim Dt^1$) with an arbitrary prefactor chosen to intercept the data at the end of the process. This is shown as the dotted line in Fig. 5a (note that this line should be taken to guide the eye, as the instantaneous diffusion coefficient of the polymer depends upon the fractional amount of monomers set in motion by the applied force).

Examining Fig. 5a, the fluctuations start very near the deterministic result. However, the deterministic line saturates near $t/\langle t\rangle \approx 0.08$. This corresponds to the time where the tension front reaches the last monomer. Correspondingly, the fluctuations in the simulations begin to level off. In general, it is assumed that the motion of a translocating polymer follows a power law $\langle \Delta s^2\rangle \sim t^\beta$ with β close to unity [39]. Fluctuations arising from thermal noise should eventually dominate over the effect from initial conformations which saturate before the translocation time $\langle\tau\rangle$. We find that the linear $\langle \Delta s^2\rangle \sim t^1$ line guides closely follows the data for long times.

The family of curves presented in Fig. 5b is for varying pre-stretching forces. As expected, polymers with higher pre-stretching exhibit a lower amount of fluctuations: as $\hat{F}_{\mathrm{s}}$ is increased the curves shift downwards. The curves differ not only in the absolute amount of fluctuations but also in the rate at which these fluctuations grow. As expected in the final—Brownian noise only—stage of translocation, the fluctuations appear to follow the expected diffusive power law. In situations of low pre-stretching, this diffusive regime is short-lived since the onset occurs very close to the end of the translocation process. In the opposing limit of high pre-stretching, the diffusive regime completely dominates. In early times however, fluctuations arise from both the conformational noise (molecular individualism) and Brownian diffusion. Hence, pre-stretched polymers start with a low variance in s that increases approximately linearly with t while polymers at $\hat{F}_{\mathrm{s}} = 0$ start with a large variance in s and essentially do not exhibit the diffusive regime. In between these extremes, a cross-over from initial conformation to thermal noise can be observed. The crossover (close to the tension-propagation time) depends upon the amount of pre-stretching, consistent with references [27, 40].

D. Scaled variations

It is convenient to calculate the standard deviation σ_τ normalized by the mean translocation time $\langle\tau\rangle$. This quantity, which reflects the relative width of the distributions, is known as the coefficient of variation and is given by

$$c_v = \frac{\sigma_\tau}{\langle\tau\rangle}. \tag{4}$$

These values are plotted in Fig. 4 b. Now, almost without exception, the normalized distribution width decreases with increasing $\hat{F}_{\mathrm{s}}$ thus indicating that reducing the range of initial conformations does reduce the uncertainty in the measured translocation time, even though translocation takes longer.

To quantify these effects further, the percent difference in the coefficient of variation χ_v is calculated as the difference between c_v with no stretching force and c_v at the highest stretching force normalized by c_v at $\hat{F}_{\mathrm{s}} = 0$:

$$\chi_v \equiv \frac{c_v(\hat{F}_{\mathrm{s}} = 0) - c_v(\hat{F}_{\mathrm{s}} = 25)}{c_v(\hat{F}_{\mathrm{s}} = 0)} \times 100. \tag{5}$$

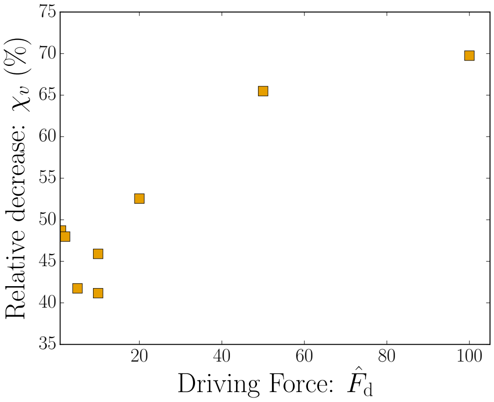


FIG. 6: (Color online) Plot of the percent decrease on the coefficient of variation c_v as a function of the driving force $\hat{F}_\mathrm{d}$.

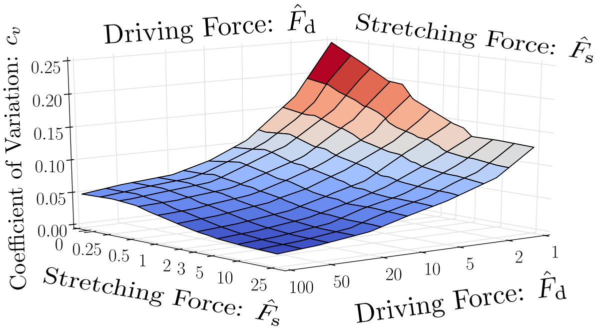


FIG. 7: (Color online) Phase diagram of the coefficient of variation c_v as a function of the two variables that control for the two noise sources; pre-stretching force $\hat{F}_\mathrm{s}$ and the driving force $\hat{F}_\mathrm{d}$. Note the nonlinear layout of both force axes.

The decrease in the variance of the translocation times is strongly dependent on the driving regime $\hat{F}_\mathrm{d}$. For $\hat{F}_\mathrm{d} = 1 - 10$ for instance, a decrease of $40 - 50\%$ is observed. At these low driving forces, the relative decrease actually decreases slightly with increasing driving force. Above $\hat{F}_\mathrm{d} = 10$, the relative decrease increases significantly with increased driving force. For $\hat{F}_\mathrm{d} = 100$, the relative width of the distribution decreases by 70% between $\hat{F}_\mathrm{s} = 0$ and $\hat{F}_\mathrm{s} = 25$. The data is saturating at large stretching forces. Note that there is a slight discrepancy between the $k_\mathrm{B}T = 0.1$ and $k_\mathrm{B}T = 0.01$ cases which do not quite overlap at $\hat{F}_\mathrm{d} = 10$.

Figure 7 displays this same data as a phase diagram in $\hat{F}_\mathrm{s}$ and $\hat{F}_\mathrm{d}$. One can see that both increasing the effective driving force (by increasing the field or decreasing the temperature) and stretching the polymer yield improved (lower) coefficients of variation. The largest effect is observed when both of these factors are employed simultaneously. However, the region near the lowest c_v is quite flat indicating that both effects saturate and thus nearly optimal results are obtained once a sufficient strong stretching force is coupled with a sufficiently large Péclet number.

IV. CONCLUSIONS

In this manuscript, we constructed a simulation model to study translocation such that two sources of noise (conformational and Brownian) can be modulated via two control parameters. This was achieved by a stretching-pulling force scenario in which a force of F_{s} is applied to the first monomer and $-F_{\mathrm{s}}$ is applied to the last monomer to stretch out the polymer. These two forces keep the polymer stretched while an additional force F_{d} is applied to the first monomer which pulls the polymer through the pore into the *trans*-region.

We used $k_{\mathrm{B}}T$ to control for Brownian noise and F_{s} for initial conformations across a range of translocation Péclet numbers by varying F_{d} and $k_{\mathrm{B}}T$.

We found that prestretching the polymer has two significant benefits: not only does the variance of τ decrease, but the mean translocation time increases. Hence, the process is slowed down and the precision of the measured times are increased; both of these are beneficial for DNA sizing and sequencing technologies. The measured benefits are found to strongly depend on the ratio $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T$ since high amounts of diffusion can overwhelm the benefits of limiting the range of initial polymer conformations.

Although the present implementation is difficult to realize experimentally, these results suggest that pre-stretching DNA via some mechanism could be very beneficial for nanopore-based sorting and sequencing applications. Current work focuses on experimentally viable methods of prestretching DNA that preserve these beneficial effects.

V. ACKNOWLEDGEMENTS

Simulations were performed using the ESPResSo package [41] on the SHARCNET computer system (www.sharcnet.ca) using VMD [42] for visualization. This work was funded by NSERC (RGPIN/046434-2013) and the University of Ottawa. HWdH gratefully acknowledges funding from NSERC via a Discovery Grant (2014-06091)

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, 2002), 4th ed.

[2] M. Muthukumar, *Polymer Translocation* (CRC Press, 2011).

[3] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, Nano Lett **5**, 1193 (2005).

[4] P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, and D. Branton, Nano Lett **4**, 2293 (2004).

[5] A. Ivankin, R. Y. Henley, J. Larkin, S. Carson, M. L. Toscano, and M. Wanunu, ACS nano **8**, 10774 (2014).

[6] C. A. Merchant, K. Healy, M. Wanunu, V. Ray, N. Peterman, J. Bartel, M. D. Fischbein, K. Venta, Z. Luo, A. C. Johnson, et al., Nano letters **10**, 2915 (2010).

[7] K.-H. Paik, Y. Liu, V. Tabard-Cossa, M. J. Waugh, D. E. Huber, J. Provine, R. T. Howe, R. W. Dutton, and R. W. Davis, ACS nano **6**, 6767 (2012).

[8] R. Tahvildari, E. Beamish, V. Tabard-Cossa, and M. Godin, Lab on a chip **15**, 1407 (2015).

[9] W. Sung and P. J. Park, Phys Rev Lett **77**, 783 (1996).

[10] D. Mondal and M. Muthukumar, J Chem Phys **144**, 144901 (2016).

[11] T. Saito and T. Sakaue, Phys Rev E **85**, 061803 (2012).

[12] T. Saito and T. Sakaue, Eur Phys J E **34**, 1 (2011).

[13] T. Sakaue, Phys Rev E **76**, 021803 (2007).

[14] T. Sakaue, Phys Rev E **81**, 041808 (2010).

[15] T. Saito and T. Sakaue, Phys Rev E **88**, 042606 (2013).

[16] P. Rowghanian and A. Grosberg, J Phys Chem B **115**, 14127 (2011).

[17] P. Rowghanian and A. Y. Grosberg, Phys Rev E **86**, 011803 (2012).

[18] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung, Phys Rev E **85**, 051803 (2012).

[19] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung, J Chem Phys **137**, 085101 (2012).

[20] I. Huopaniemi, K. Luo, T. Ala-Nissila, and S.-C. Ying, J Chem Phys **125**, 124901 (2006).

[21] I. Huopaniemi, K. Luo, T. Ala-Nissila, and S.-C. Ying, Phys Rev E **75**, 061912 (2007).

[22] C. Forrey and M. Muthukumar, J Chem Phys **127**, 015102 (2007).

[23] H. Katkar and M. Muthukumar, J Chem Phys **140**, 135102 (2014).

[24] J. M. Polson, J Chem Phys **142**, 174903 (2015).

[25] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, Biophys J **101**, 70 (2011).

[26] J. Sarabadani, T. Ikonen, and T. Ala-Nissila, J Chem Phys **141**, 214907 (2014).

[27] D. Sean, H. W. de Haan, and G. W. Slater, Electrophoresis **36**, 682 (2015), ISSN 1522-2683.

[28] D. Sean and G. Slater, Journal of Chemical Physics **Under review** (2016).

[29] U. Keyser, J. Van der Does, C. Dekker, and N. Dekker, Review of Scientific Instruments **77**, 105105 (2006).

[30] A. Sischka, A. Spiering, M. Khaksar, M. Laxa, J. König, K.-J. Dietz, and D. Anselmetti, Journal of Physics: Condensed Matter **22**, 454121 (2010).

[31] V. Tabard-Cossa, M. Wiggin, D. Trivedi, N. N. Jetha, J. R. Dwyer, and A. Marziali, ACS nano **3**, 3009 (2009).

[32] S. Ollila, K. Luo, T. Ala-Nissila, and S.-C. Ying, The European Physical Journal E **28**, 385 (2009).

[33] V. V. Lehtola, R. P. Linna, and K. Kaski, Phys Rev E **78**, 061803 (2008).

[34] G. Slater, C. Holm, M. Chubynsky, H. De Haan, A. Dubé, K. Grass, O. Hickey, C. Kingsburry, D. Sean, T. Shendruk, et al., Electrophoresis **30**, 792 (2009).

[35] J. D. Weeks, D. Chandler, and H. C. Andersen, J Chem Phys **54**, 5237 (1971).

[36] G. Grest and K. Kremer, Phys Rev A **33**, 3628 (1986).

[37] H. W. de Haan, D. Sean, and G. W. Slater, Phys Rev E **91**, 022601 (2015).

[38] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, Phys Rev E **87**, 032147 (2013).

[39] J. Dubbeldam, A. Milchev, V. Rostiashvili, and T. A. Vilgis, EPL (Europhysics Letters) **79**, 18002 (2007).

[40] P. Rowghanian and A. Y. Grosberg, Phys Rev E **86**, 011803 (2012).

[41] H.-J. Limbach, A. Arnold, B. A. Mann, and C. Holm, Comput Phys Commun **174**, 704 (2006).

[42] W. Humphrey, A. Dalke, and K. Schulten, J Molec Graphics **14**, 33 (1996).

6

Pulling a prestretched polymer through a nanopore: comparison with the Tension-Propagation theory of translocation

David Sean, Hendrick W. de Haan, Gary W. Slater

This is a short theoretical addendum to the manuscript found in the previous chapter. A high-field tension-propagation model is derived to include the effects of end-pulling a pre-stretched polymer. I plan on continuing this work and to submit the results in a manuscript for peer review. What follows is mostly original work with advice from Hendrick and Gary.

Pulling a prestretched polymer through a nanopore: comparison with the Tension-Propagation theory of translocation

David Sean and Gary W. Slater
Physics Department, University of Ottawa,
Ottawa, Ontario, Canada, K1N 6N5

Hendrick W. de Haan
Faculty of Science, University of Ontario Institute of Technology,
Oshawa, Ontario, Canada, L1H 7K4

(Dated: December 23, 2016)

INTRODUCTION

This chapter serves as a short theoretical addendum to the previous chapter and will eventually be transformed into a manuscript to be submitted for peer reviewed publication. A theoretical model for describing the conformations of a tethered polymer under pre-stretching forces is developed based on the Marko-Siggia force-extension relationship and blob theory. Furthermore, this model is used in a Tension-Propagation (TP) model for polymer translocation driven by pulling on an end.

Most TP models neglect the crowding of monomers on the *trans*-side and focus on the drag obtained when *cis*-monomers are pulled. Translocation by end-pulling reduces the number of *trans* monomers near the nanopore exit. Thus, the impact of neglecting the crowding contribution of the *trans* monomers is reduced in the end-pulling setup. This provides a more direct comparison between the TP theory and computer simulations.

MODEL

We will consider the *strongly-stretched* regime of tension-propagation [1–5]. We can thus neglect thermal fluctuations and the monomer dynamics are considered to be purely deterministic on the time scale of translocation. However, as described by Satio and Sakaue, there remains a stochastic element which stems from the initial polymer conformations at the beginning of the translocation event [4]. We demonstrated in the previous chapter that modulating the initial conformations via a pre-stretching force has a higher impact in this regime.

We derive here a TP picture of translocation via end-pulling following the approach proposed by Saito and Sakaue [4]. We model the polymer as a tethered self-avoiding walk with the first monomer inside the nanopore. As translocation takes place, the monomers are sequentially displaced from the *cis* to the *trans* side. The translocation coordinate s is given by the monomer number inside the pore. Since we want to restrict our model to single-file translocation, this also coincides with the total number of monomers on the *trans*-side. For the force setup considered here, polymer translocation is accomplished by pulling on the first monomer via a driving force $F_{\mathrm{d}}\hat{x}$, as shown in Fig. 1. This can help reduce *trans* crowding and will significantly increase the total number of dragged monomers as opposed to when

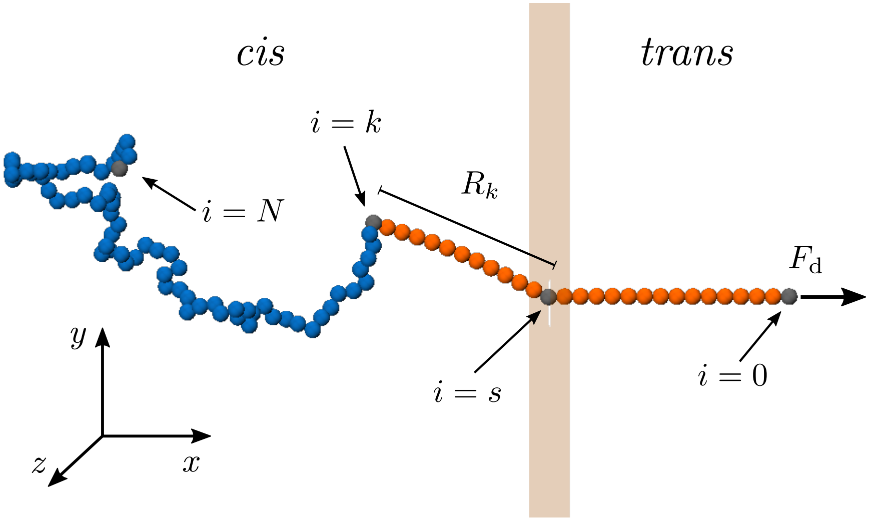


FIG. 1. Schematic for driven translocation of a polymer with the driving force $F_\mathrm{d}\hat{x}$ applied to the first monomer $i = 0$. The translocation coordinate s is given by the monomer residing in the pore. The section under tension on the *cis*-side, which is of length R_k, is bounded by the monomer in the pore (s) and the monomer at the tension front (k). The tension front grows until it reaches the last monomer, $i = N = 100$.

the force is applied to the monomer in the pore.

For simplicity, we will consider the deterministic situation where the Brownian motion of the monomers are neglected, thus *cis* monomers are considered immobile until the tension due to the driving force in the pore propagates and initiates motion. With this picture in mind, we can define the location of the tension front as the lowest monomer index which is still in its initial position. The cartoon in Fig 1 shows a highlighted polymer segment (on both *cis* and *trans* sides) which is under tension up to the kth monomer. In the strongly-stretched regime, the taut segment forms a single-file stem between monomers $i = k$ and the end $i = 0$.

There are thus k monomers moving due to the applied force F_d at all times. On the *cis*-side only however, there are $k - s$ such that

$$R_k/b = \tilde{R}_k = k - s. \tag{1}$$

where b is the monomer size. The non-dimensionalised distance $\tilde{R}_k$ holds information pertaining to the initial conformation ($\tilde{R}_k$ can be interpreted as the distance between the k^{th}

monomer and the nanopore). Its time derivative is given as

$$\frac{d\tilde{R}_k}{dt} = \frac{dk}{dt} - \frac{ds}{dt}. \tag{2}$$

We note that the last term gives the translocation rate ds/dt.

The translocation process is separated in two time stages: the Tension-Propagation stage with a duration τ_{TP} where the tension front grows until it reaches the last monomer ($k = N$), followed by the Post-Propagation stage, with a duration τ_{PP} where the tension front stops growing ($\frac{dk}{dt} = 0$) and the remainder of the polymer is sucked in the pore as a single body of length $R_N + sb$.

In the over-damped regime the applied force is balanced by the viscous force on the k moving monomers. This yields a terminal velocity

$$v = \frac{F_{\mathrm{d}}}{\zeta k}. \tag{3}$$

Since these monomers are displaced as a single group, which includes the monomer inside the pore s, the translocation rate ds/dt is essentially given by v/b. Substituting this into Eq. 2 allows us to separate the variables to obtain

$$\frac{F_{\mathrm{d}}}{\zeta b} dt = k dk - k d\tilde{R}_k. \tag{4}$$

Integrating over $k \in (0, N)$ and $\tilde{R}_k \in (0, \tilde{R}_N)$, where $\tilde{R}_N$ is given by the initial position of the N^{th}—and last— monomer, to find τ_{TP} gives

$$\frac{F_{\mathrm{d}}}{\zeta b} \tau_{\mathrm{TP}} = \frac{1}{2} N^2 - \int_0^{\tilde{R}_N} k d\tilde{R}_k. \tag{5}$$

In the post-propagation step, $\frac{dk}{dt} = 0$, $k = N$, and (the shrinking) $\tilde{R}_k$ runs over $\tilde{R}_k \in (\tilde{R}_N, 0)$ which gives

$$\frac{F_{\mathrm{d}}}{\zeta b} \tau_{\mathrm{PP}} = -N \int_{\tilde{R}_N}^0 d\tilde{R}_k = N\tilde{R}_N. \tag{6}$$

Combining the duration of the two steps gives a total translocation time

$$\tau = \tau_{\mathrm{TP}} + \tau_{\mathrm{PP}} = \frac{\zeta b}{F_{\mathrm{d}}} \left(\frac{1}{2} N^2 + N\tilde{R}_N - \int_0^{\tilde{R}_N} k d\tilde{R}_k \right). \tag{7}$$

We note here that the natural time unit is $\zeta b / F_{\mathrm{d}}$. In the highly driven limit, changing the driving force F_{d} results in a trivial scaling of the translocation time since $\tau \sim 1/F_{\mathrm{d}}$.

The translocation time τ is controlled entirely by the initial conformation. In the discrete simulation model this is the set $\{R_k\}$. In the following section, we propose an analytical form to describe the mean initial conditions as function of the stretching force F_s using a modified Marko-Siggia model.

In the limit of a polymer having a conformation forming a straight rod aligned with the nanopore axis (which corresponds to the highest extension $F_\mathrm{s} \to \infty$), we can use $R_k = kb$. For this extreme case, the tension immediately reaches the tip ($\tau_\mathrm{TP} = 0$), while the translocation time is $\tau = \tau_\mathrm{TP} = N^2\zeta b/F_\mathrm{d}$. Interestingly, this is also the scaling found for pore driven translocation of a polymer initialized in a rod-like conformation [6] since this limit does not depend on which monomer the force is applied to (single-body movement). What thus remains to be determined is a general form for $\{R_k\}$ for the case of a tethered polymer subject to an arbitrary stretching force F_s.

Simulations

Coarse-grained Langevin Dynamics (LD) simulations are performed in order to test Eq. 7. These simulations will be used not only to compute translocation times, but also to generate an ensemble of initial conformations. This simulation approach is identical to that of the preceding chapter. The stretching force used is $F_\mathrm{s} = \hat{F}_\mathrm{s} k_\mathrm{B}T/\sigma$, such that the deformation from the dimensionless force $\hat{F}_\mathrm{s}$ is independent of temperature.

We supplement the investigation with additional simulations of deterministic translocations using $k_\mathrm{B}T = 0$. For these $k_\mathrm{B}T = 0$ cases, the initial conformations under pre-stretching is first obtained at nonzero values of $k_\mathrm{B}T$; after the equilibration, we set $k_\mathrm{B}T = 0$ as well as $F_\mathrm{s} = 0$ such that the initial conformation remains unaltered.

Changing driving regimes is not of primary interest here. We thus employ a single driving force of $F_\mathrm{d} = 1\epsilon_\mathrm{LJ}/\sigma$. The results are presented using the time unit $\tau_\mathrm{sim} = \zeta\sigma/F_\mathrm{d}$.

RESULTS

Translocation

We first test the TP theory using the discrete set $\{R_k\}$ as obtained from the LD simulations at ($t = 0$). For this, we compare the predictions of Eq. 7 to the complete LD

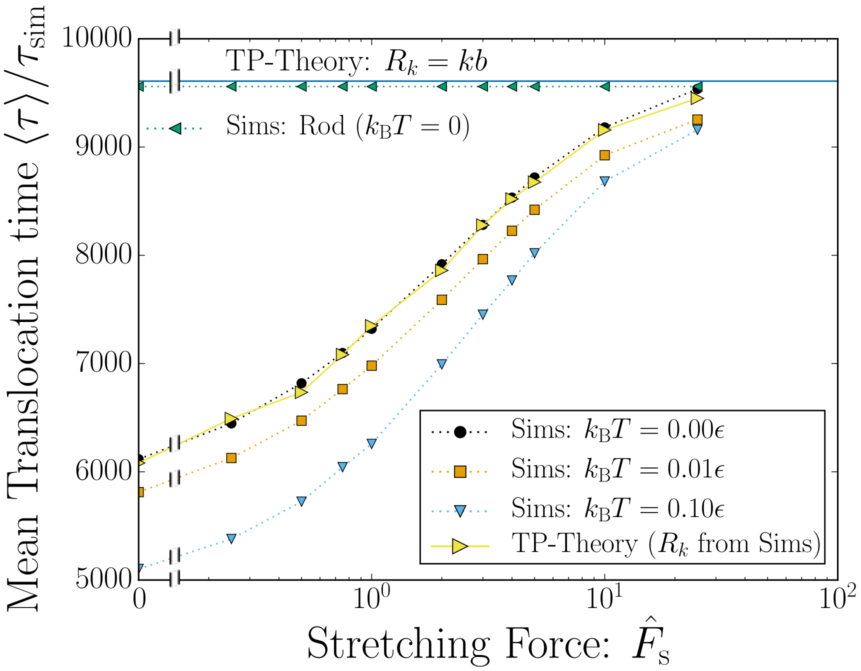


FIG. 2. (Color online) For $N = 100$, we show the average translocation time obtained from the LD simulations (dotted lines) and the results of TP-theory (full lines) using the initial conformations $\{R_k\}$ from the same simulations. The rod-like limit is both simulated and evaluated with TP theory using $R_k = kb$.

simulations. Figure 2 shows the average translocation time for different pre-stretching forces $\hat{F}_s$. After the polymer is sufficiently warmed-up, the initial conformation is saved for later TP analysis (which needs $\{R_k\}$), and the translocation is fully simulated using LD. A fully deterministic $k_BT = 0$ case is added here to illustrate the highly driven limit and verify the Tension-Propagation theory.

Using the saved conformations, individual monomer position sets $\{R_k\}$ are computed and the theoretical translocation time is evaluated using Eq. 7. The corresponding values shown in Fig. 2 agree quantitatively with the $k_BT = 0$ limit.

Both complete simulations and predictions from TP theory are shown for the limiting "rod" case where we used $R_k = kb$ (with $b = 0.96\sigma$) instead of the simulated initial monomer positions. Both methods agree for this trivial limiting case.

Since we use the single value of $F_d = 1\epsilon_{LJ}/\sigma$, the $k_BT = 0.1\epsilon_{LJ}$ and $k_BT = 0.01\epsilon_{LJ}$ simulation data sets correspond to different scaled driving forces $\hat{F}_d = F_d\sigma/k_BT$ and contain the temperature effects discussed in Chapter 5. Increasing the temperature results in decreasing

$\langle \tau \rangle$ for all stretching forces $\hat{F}_{\rm s}$. Inspection of both the full simulations and the TP-theory results in Fig. 2 confirms the validity of TP-theory in the $F_{\rm d} \gg k_{\rm B}T/\sigma$ limit.

Although the simulation results and the TP theory agree, the initial conformations must still be generated via simulations. We now turn to modelling R_k in order to obtain theoretical curves without any explicit simulation input.

Prestretching a tethered polymer: Finding the function $R_k(\xi)$

We first examine the case without a pre-stretching force ($F_{\rm s} = 0$) to obtain the prefactor A_0 in the presence of the wall (see Eq. 8 below). An ensemble of 8000 conformations of a polymer with the first monomer fixed inside the nanopore are generated in the simulations. In the situation where no external force is applied to the untethered polymer end, we expect self avoiding walk statistics of the form

$$R_k = A_0 b k^{\nu}, \tag{8}$$

where $\nu = 0.588$ is the size exponent and A_0 is a model-dependent prefactor which includes the effects of tethering the polymer in half-space. A fit to our simulation data gives $A_0 = 1.545 \pm 0.003$ as shown in Fig. 3.

Under the effect of a stretching force $\hat{F}_{\rm s}$, we take the conformation to be well represented by a chain of blobs of size ξ with g monomers per blob. In this representation, the initial distance R_k scales linearly with monomer index k, thus $R_k \sim (\xi/g)\, k$. However, for length scales less than the blob size, we expect the aforementioned self avoiding statistics $R_k = A_0 b k^{\nu}$ to hold.

In order to combine these two, we find that our data can be well represented by a piecewise fit of the form

$$R_k(\xi) = \begin{cases} A_0 b k^{\nu} & \text{for } k < k^* \\ \left(\frac{\xi}{b}\right)^{1-1/\nu} \sigma k + B_0 & \text{for } k \geq k^* \end{cases} \tag{9}$$

where B_0 is a constant chosen such that the function is continuous at k^*, the crossover monomer index. A naïve crossover k^* would be $k^* = g$, the number of monomers per blob, such that the crossover occurs after the first blob is populated. However this choice gives a $R_k(\xi)$ relation which is not smooth: the derivative is discontinuous at $k = k^*$. Matching the

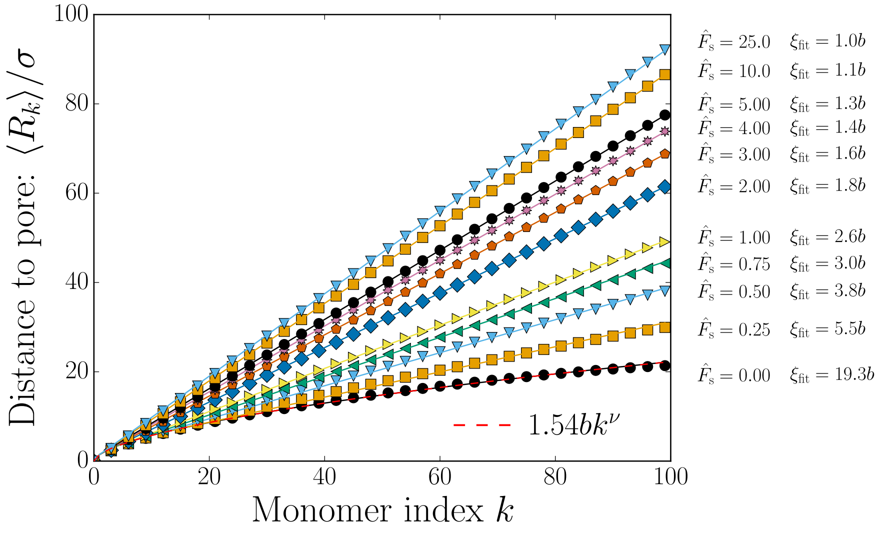


FIG. 3. (Color online) Fit of the mean polymer extension $\langle R_k \rangle$ as a function of the monomer index k for a tethered polymer being stretched by a force $\hat{F}_{\rm s}$. The last $\hat{F}_{\rm s} = 0$ case is also fitted (red dashes) with Eq. 8. The fitted values of the blob size ξ are shown in the legend. The symbols are simulation data and the fits are shown as lines.

derivative with respect to k of the two pieces yields:

$$k^* = \left(\frac{b}{\xi}\right)^{(1-\nu)/\nu} (A_0 b \nu)^{1/(1-\nu)}\,, \tag{10}$$

which results in a smooth and continuous crossover when

$$B_0 = A_0 b k^{*\nu} - \left(\frac{b}{\xi}\right)^{1/\nu} \xi k^* \tag{11}$$

is chosen.

Figure 3 shows the results of the fits using the blob size ξ as a free parameter ($A_0 = 1.545$ and $\nu = 0.588$ are kept fixed). We include in this figure the case of $\hat{F}_{\rm s} = 0$ fitted with the form $R_k = A_0 b k^{\nu}$—note that the piece-wise fit is identical since $k^* = N$. The fitted values of $\xi_{\rm fit}$ are useless to generate the $R_k(\xi)$ profiles to be used in the TP theory since the goal is to remove the need for the simulations. However, our data suggest that the piece-wise construction of Eq. 9 can now be used with a fair degree of confidence in order to provide the needed $\{R_k\}$. This effectively reduces the complexity of the input parameter from the N-sized set $\{R_k\}$ to a single value ξ. In order to obtain an estimate for ξ as a function of the pre-stretching force $\hat{F}_{\rm s}$, we now turn to a modified Marko-Siggia force-extension relation.

Using the Marko-Siggia formular for ξ

When subject to two opposing forces of magnitude F_s, the polymer end-to-end distance R_N initially increases as if it were a harmonic spring for low forces before an asymptotic transition towards the finite length L_c at high stretching. For DNA, which is semiflexible, these two behaviours are often modelled using the force-extension relation proposed by Marko and Siggia [7]:

$$F_\mathrm{MS} = \frac{1}{4}\left(1 - \frac{R_N}{L_\mathrm{c}}\right)^{-2} - \frac{1}{4} + \frac{R_N}{L_\mathrm{c}}. \tag{12}$$

To model the pre-stretched conformation of a polymer where the first monomer resides in the pore, the Marko-Siggia relationship cannot be taken at face value here since the presence of the membrane imposes the condition that the other monomers can only be found in half the available space. Clearly, Eq. 12 predicts $R_N = 0$ at zero stretching forces which is incompatible with the situation of a polymer tethered unto a wall. In the limit of high stretching forces however, the effect of the wall should become negligible since the majority of monomers will not sample the geometric restrictions imposed by the wall. We thus expect Eq. 12 to become applicable in this limit.

The Marko-Siggia relation can be generalized in slit-like geometries using an *effective* dimensionality [8]. To the best of my knowledge no such theoretical generalization has been made for a polymer tethered to a wall. Interestingly, experimental validation of the Marko-Siggia relation has been realized by DNA stretching where one end was tethered to a wall and the other was pulled by a magnetic bead [7, 9].

We propose here a minor addition of a constant R_0 such that the nonzero value of R_N at $F_\mathrm{s} = 0$ can be captured. In principle, the correction term itself should vanish with increasing stretching force $\hat{F}_\mathrm{s}$ since the monomers forming an aligned "string of blobs" no longer interact with the wall. However, the correction term R_0 will become relatively small in comparison with the extension R_N when F_s is sufficiently large. Thus we approximate R_0 as a *constant* with the assumption that its contribution will be dwarfed in cases where it should vanish. Including the shift into Eq. 12 gives

$$F'_\mathrm{MS} = \frac{1}{4}\left(1 - \frac{R_N}{L_\mathrm{c}}\right)^{-2} - \frac{1}{4} + \frac{R_N}{L_\mathrm{c}} - \frac{R_0}{L_\mathrm{c}} \tag{13}$$

which coincidentally is identical to a modification proposed in ref [10] for slit-like geometries. The equilibrium (unstretched) conformations of a tethered polymer was investigated in the

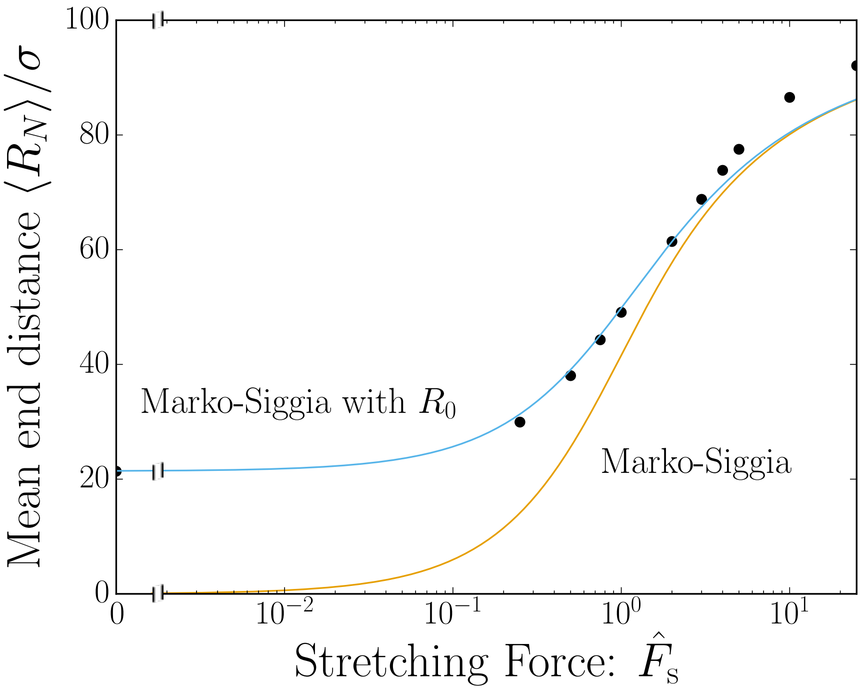


FIG. 4. Force extension curves for the tethered polymer. Simulation data are shown as symbols. The solid lines show the Marko-Siggia force-extension relation together with a version offset by R_0 to capture the non-stretched tethered polymer extension.

previous subsection, see Eq. 8. Using the A_0 value obtained, we find $R_0 \approx 22.3b$ for our $N = 100$ simulation polymer.

In Fig. 4, we show simulated force-extension data resulting from having the first monomer confined to the nanopore entrance. The Marko-Siggia Eq. 12 equation fails to capture the finite extension at vanishing stretching forces F_s. Adding a non-zero extension R_0 effectively translates the curve vertically.

From the modified Marko-Siggia relation, one can obtain R_N as a function of the prestretching force for a tethered polymer. Since our piece-wise Eq. 9 permits the full profile *including* the last point $R_{k=N}$, we can effectively map out $R_k(\hat{F}_\mathrm{s})$. The simplest way is to iterate twice. Although Eq. 13 cannot be analytically inverted, one can numerically obtain the position of the last monomer R_N for a given force $\hat{F}_\mathrm{s}$. With this information, we can turn to Eq. 9, which predicts the full profile and find the value of ξ such that the last point is the same as was found with the modified Marko-Siggia $R_{k=N}(\xi) = R_N(F'_\mathrm{MS})$.

Once ξ is found, we can use Eq. 9 again to generate the full $R_k(\xi)$ profile (by construc-

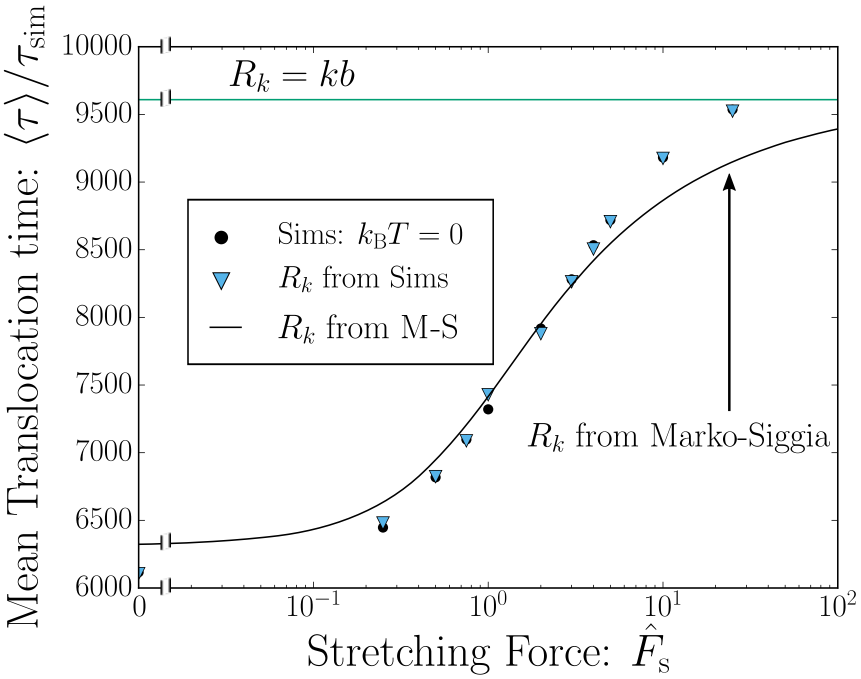


FIG. 5. The mean translocation time as a function of $\hat{F}_\mathrm{s}$. Deterministic simulations ($k_\mathrm{B}T = 0$) are shown almost completely overlapping with the TP theory using the same initial conformation $\{R_k\}$. The TP prediction is also obtained using a smooth function of R_k is obtained using the last point R_N as predicted by a modified Marko-Siggia (M-S) force-extension relation.

tion, the last point of $\{R_k\}$ is now identical to that obtained by the modified Marko-Siggia relation).

The full R_k can then be used in the integral of Eq. 7. This ultimately permits us to directly plot the mean translocation time as a function of $\hat{F}_\mathrm{s}$, as shown in Fig. 5.

The deterministic simulation is re-plotted here as well as the TP predictions using the initial conformations from the simulations ($\{R_k\}$). The asymptotic limit $R_k = kb$ is also included for reference. As Fig 5 shows, the theoretical curve obtained via the modified Marko-Siggia relation qualitatively (and to some degree quantitatively) agrees with the simulation data. Here however, the data is now a continuous function of $\hat{F}_\mathrm{s}$.

There are some high field disagreements between the mean translocation time $\langle\tau\rangle$ predicted by this theoretical model and the simulation data. We can work backwards in order to target the approximation affecting the discrepancies. The piece-wise function to find $R_k(\xi)$ in Eq. 9 is probably not the weak point considering the remarkable (albeit fitted) agreement shown in Fig. 3 (including the last point). The Marko-Siggia modifications of Eq. 13, however, also shares disagreements at high force, see Fig. 4. The Marko-Siggia relation is derived from the wormlike chain model which considers the polymer as a continuous curve. At the stretching forces such that the blob size approaches our monomer size, the discrete nature of the simulation model may be responsible for this disagreement. We are currently exploring this hypothesis as well as ways in which a force-extension relation can better capture the simulation results.

CONCLUSION

Following the work of Saito and Sakaue [4], a tension-propagation model for polymer translocation in the highly-driven limit was developed for end-pulling translocation. We investigated the effect of modulating the polymer conformations prior to translocation by the use of a pre-stretching force $\hat{F}_\mathrm{s}$.

The high-field limit of this model results in a translocation time that scales inversely with the driving force F_d (see Eq. 7). The translocation time is obtained by integrating a term $k(dR_k/dk)dk$ over $k \in (0, N)$ where R_k is the distance between the k^{th} monomer and the nanopore as measured in the initial conformation. In order to evaluate this integral, Langevin dynamics were used to generate the conformations ($\{R_k\}$), as well as to integrate the complete translocation process in time. In the limit $k_\mathrm{B}T \to 0$ (equivalent to the high-field limit), the data from the LD simulations approach the TP theory predictions. The deterministic simulations show quantitative agreement with the TP theory that uses $\{R_k\}$ from the simulations.

We proposed a continuous interpolation to obtain $R_k(\xi)$ as a function of a single parameter, blob size ξ, or equivalently the position of the last monomer $R_k(R_N)$. To obtain the latter, a Marko-Siggia model was generalized to account for the repulsion of the wall at zero stretching force.

We thus developed a theoretical model to predict the translocation time of pre-stretched polymers where the translocation is driven by end-pulling, as investigated by simulations in Chapter 5. The model agrees with the computer simulations in the limit of high driving forces.

---

[1] T. Sakaue, Phys Rev E **76**, 021803 (2007).

[2] T. Sakaue, Phys Rev E **81**, 041808 (2010).

[3] T. Saito and T. Sakaue, Eur Phys J E **34**, 1 (2011).

[4] T. Saito and T. Sakaue, Phys Rev E **85**, 061803 (2012).

[5] T. Saito and T. Sakaue, Phys Rev E **88**, 042606 (2013).

[6] V. V. Lehtola, R. P. Linna, and K. Kaski, Phys Rev E **78**, 061803 (2008).

[7] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).

[8] H. W. de Haan and T. N. Shendruk, ACS Macro Letters **4**, 632 (2015).

[9] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).

[10] Y.-L. Chen, P.-k. Lin, and C.-F. Chou, Macromolecules **43**, 10204 (2010).

7

Langevin Dynamics simulations of driven polymer translocation into a crosslinked gel

David Sean, Gary W. Slater

Accepted (in press) in *Electrophoresis*, special issue on Fundamentals.

Langevin Dynamics Simulations of Driven Polymer Translocation into a Crosslinked Gel

David Sean, Gary W. Slater

Department of Physics, University of Ottawa,
150 Louis-Pasteur, Ottawa, Ontario, Canada, K1N 6N5

December 25, 2016

Corresponding author:
Prof. Gary W. Slater
email: gary.slater@uOttawa.ca
fax: +1-613-562-5190

Keywords:

Polymer Translocation, Gels, Nanopore, DNA Sizing, DNA Sequencing

Running title:

Simulations of Driven Polymer Translocation into a Crosslinked Gel

Abbreviations:

TP	Tension-Propagation
LD	Langevin Dynamics
WCA	Weeks-Chandler-Andersen potential
FENE	Finitely Extensible Nonlinear Elastic
IMFPT	Incremental Mean First Passage Time

Abstract

We investigate the dynamics of driving a polyelectrolyte such as DNA though a nanopore and into a crosslinked gel. Placing the gel on the *trans*-side of the nanopore can increase the translocation time while not negatively affecting the capture rates. This setup thus combines the mechanics of gel electrophoresis with nanopore translocation. However, contrary to typical gel electrophoresis scenarios, the effect of the field is localized in the immediate vicinity of the nanopore and becomes negligible inside the gel matrix. We thus investigate the process by which a semiflexible polymer can be *pushed* into a gel matrix via a localized field and we describe how the dynamics of gel penetration depends upon the field intensity, polymer stiffness and gel pore size. Our simulation results show that a semiflexible polymer enters the gel region with two distinct mechanisms depending upon the ratio between the bending length scale and the gel pore size. In both regimes the gel fibres cause a net increase of the mean translocation time. Interestingly, the translocation rate is found to be constant (a potentially useful feature for many applications) during the predominant part of the translocation process when the polymer is stiff over a length scale comparable to the gel pore size.

1 Introduction

The forced translocation of DNA from one side of a membrane to the other via a small nanoscopic hole, such that the bases pass in a sequential manner, has been proposed as a new approach to DNA sequencing; not surprisingly, this idea has also motivated numerous fundamental studies of driven polymer translocation over the last decade [1, 2, 3]. This transport process is also of biological interest. Small nuclear pore complexes create a passage through the nuclear membrane for the transit of small molecules while selectively prohibiting larger entities from entering via diffusion [4]. Understanding the active transport of the larger protein and RNA molecules through such pores has thus motivated theoretical and computational studies of nanopore translocation in the presence of obstacles or crowding agents [5, 6, 7]. Recent experiments of nanopore DNA translocation in crowded environments include placing a nanofibre mesh [8] or an agarose gel [9] on the *cis* (or injection) side as a way to slow down translocation events.

The high electrical field intensities needed to attract diffusing DNA molecules towards the nanopore imply that the resulting translocation process is in a highly driven regime where the fast passage of bases becomes a challenge for identification [10]. Decreasing the translocation rate while not decreasing the DNA capture rate has motivated the use of agarose or polyacrylamide gels casted on the *trans* (or receiving) side [10]. General results from the theory of gel electrophoresis cannot easily be ported to this situation because of the way the electric field lines are distorted near the nanopore (these theories assume relatively homogenous field lines over the length of the DNA molecule inside the gel). In the case of nanopore translocation, the majority of the driving force is applied on to the monomers residing inside the nanopore; the latter must push against the recently translocated monomers on the *trans*-side where the effect of a gel would be to slow down the process in order to help detection. Since this situation is rather unique, it is not trivial to physically interpret results where the DNA is *pushed* inside a gel.

In this manuscript, we will focus on the fundamental question of how a polymer is pushed inside a gel matrix. Using coarse-grained Langevin-Dynamics (LD) simulations, we will explore how polymer stiffness and gel pore size influence the motion of a DNA molecule

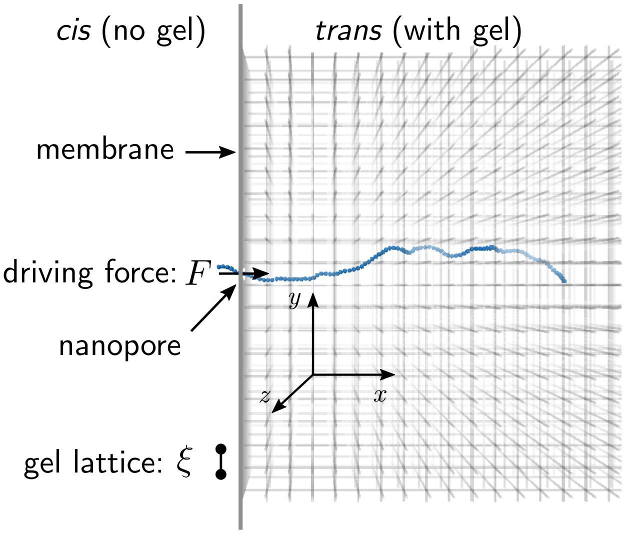


Figure 1: (Color Online) Simulation snapshot of a semiflexible polymer (with a bending length $L_{\text{Bend}} = 20\sigma$) near the end of the translocation process: the polymer moved from the gel-free *cis* side through a nanopore and into a gel matrix with a pore size $\xi = 5\sigma$. The driving force is only applied to the monomers present inside the nanopore.

pushed inside a gel. As we will show, the combined effect of the polymer stiffness and the gel fibers lead to non-trivial effects where a polymer chain may appear more rigid than expected.

2 Simulations

Similar to other coarse-grained translocation simulations, we Langevin Dynamics to integrate a generic bead spring polymer model of $N = 100$ beads with excluded volume interactions [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. The Weeks-Chandler-Andersen (WCA) potential is used to model excluded volume interactions [24]. This is simply a Lennard-Jones potential truncated at the well minimum such that it is purely repulsive [23]:

$$U_{\text{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r < r_{\text{c}} \\ 0 & \text{for } r \geq r_{\text{c}}, \end{cases} \tag{1}$$

where the nominal bead diameter σ is used as a simulation length scale, the well-depth ϵ is used as the simulation energy scale, and the cutoff distance is $r_{\text{c}} = 2^{1/6}\sigma$.

Consecutive monomers along the backbone are connected with Finitely Extensible Nonlinear Elastic (FENE) springs described by the potential

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}\, k_{\mathrm{FENE}}\, r_0^2 \ln\left[1 - \frac{r^2}{r_0^2}\right], \tag{2}$$

where the maximum extension $r_0 = 1.5\sigma$ and spring constant $k_{\mathrm{FENE}} = 30\epsilon/\sigma^2$ are used [23, 25].

The polymer stiffness is implemented by associating a bending cost to the angle between two consecutive bonds. In this study we use a harmonic potential

$$U_{\mathrm{bend}} = \kappa\,(\phi - \pi)^2\,, \tag{3}$$

where κ is the bending constant and ϕ is the angle between two consecutive bonds. We vary the degree of polymer stiffness via the bending parameter $\kappa = 0.5 - 100.0\ k_{\mathrm{B}}T$.

Free solution simulations were used to find the polymer's natural persistence length L_{p} via the decay of the autocorrelation function of the unit bond vectors $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{r}}_j$ using the expression

$$\langle \hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j \rangle = e^{-(i-j)\sigma/L_{\mathrm{p}}}. \tag{4}$$

Since the fitted values are close to the *nominal* thermal bending length scale

$$\frac{L_{\mathrm{Bend}}}{\sigma} \equiv \frac{\kappa}{k_{\mathrm{B}}T}, \tag{5}$$

we will use the latter in the rest of this article to refer to the chain's intrinsic mechanical rigidity.

The thin dividing membrane, with a nanopore of accessible diameter 2σ, is modeled via a mathematical surface. The membrane-bead interactions are described by the repulsive potential U_{WCA}; consequently, the membrane has an *effective* thickness of 1σ. The driving force $\mathbf{F} = 1\ k_{\mathrm{B}}T/\sigma\ \hat{\mathbf{x}}$ (see Fig. 1) is applied only to monomers found inside the nanopore; the extension of the latter is defined as $\pm\sigma/2$ from the centre of the membrane. The index of the monomer inside the pore also serves as the translocation coordinate s; thus the translocation process starts at $s = 1$ and ends at $s = 100$.

A Monte-Carlo method is used to pre-generate 1000 initial conformations of a self-avoiding semiflexible polymer tethered to a flat surface. To reduce the number of failed translocation events (polymers ending up on the *cis*-side of the membrane), we seed a single monomer on the *trans*-side. The LD polymer is thus initialized using the generated conformations with the first two monomers on opposite sides of the nanopore centre: one at a distance of $\sigma/2$ to the left, and one at $\sigma/2$ to the right. To avoid repeating the computationally expensive equilibration phase of the simulation, failed events are re-started with the same initial conformation.

From these initial conformations, we performed an additional autocorrelation analysis and found that the wall introduces an increase in the persistence length L_{p}. To avoid confusion, we will use the *effective* persistence length L'_{p} as a statistical measurement of the decay in the correlation function of the bond vectors in the presence of constraints, and the bending length L_{Bend} as the intrinsic structural bending scale of the polymer.

The gel is modeled by a regular cubic lattice of immobile poles (or gel fibres) placed on the *trans*-side of the membrane. The lattice spacing $\xi = 5, 6, 7\sigma$ is considered to be the pore size of the matrix. The beads interact with the thin poles via the repulsive potential U_{WCA} such that the poles have an effective diameter of σ. The fibres intersect at the cube vertices which allows to model the topology of a cross-linked gel. The nanopore is positioned to be at the centre of one of the cubic faces such that the gel fibres do not obstruct the nanopore opening, as shown in Fig. 1.

The monomer bead positions $\mathbf{r}(t)$ are integrated in time using the Langevin equation

$$m\ddot{\mathbf{r}} = \mathbf{F}^{\mathrm{C}} - \zeta\dot{\mathbf{r}} + \mathbf{F}^{\mathrm{B}}, \tag{6}$$

where m is the bead mass, $\mathbf{F}^{\mathrm{C}}$ is the sum of all conservative forces, ζ is the bead friction coefficient, and $\mathbf{F}^{\mathrm{B}}$ is an uncorrelated random noise term with zero mean and a variance of $2\zeta k_{\mathrm{B}}T/\Delta t$. Our simulation time unit is defined by the monomer diffusion time $\tau_{\mathrm{sim}} \equiv \zeta\sigma^2/k_{\mathrm{B}}T$. We use $k_{\mathrm{B}}T = 1\epsilon$, $\zeta = 1\sqrt{\epsilon/m\sigma}$ and an integration time step $\Delta t = 0.01\tau_{\mathrm{sim}}$.

The LD simulations are implemented with the ESPResSo simulation package [26]. The data are taken from an ensemble of 1000 successful translocation events.

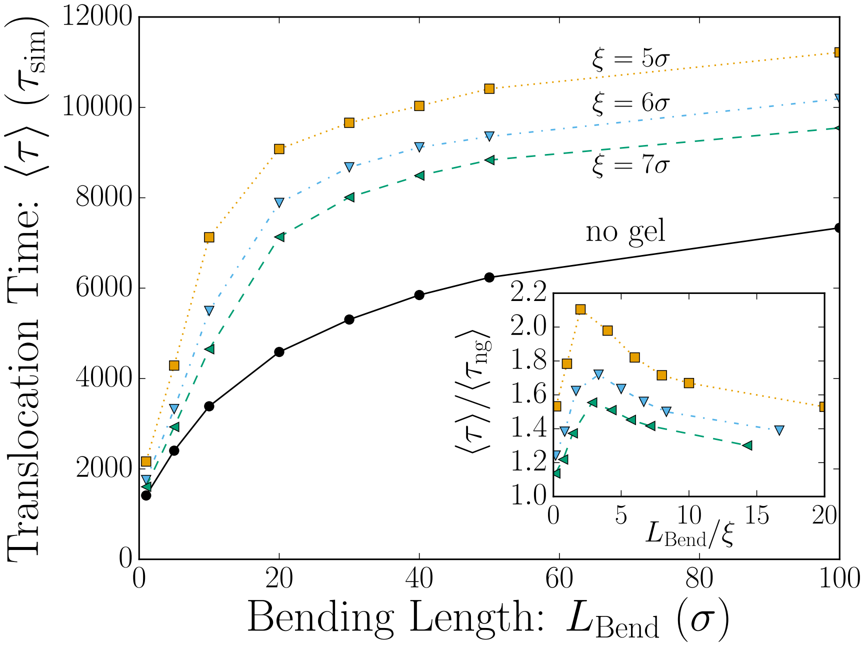


Figure 2: (Color Online) The mean translocation time as a function of the polymer bending length $L_{\rm Bend}$ for the no-gel case and for gels with pore sizes $\xi = 5, 6, 7\sigma$. Inset: Translocation times re-scaled by the no-gel value $\langle\tau_{\rm ng}\rangle$ as a function of the ratio between the bending length and the gel pore size.

3 Results

As a control, we first investigate how chain rigidity affects the translocation of a polymer when we have free solution conditions on both side of the membrane (no gel). Figure 2 shows that increasing polymer rigidity increases the mean translocation time by about a factor of four in this example (circles; solid line). Our 3D results are in qualitative agreement with 2D Brownian dynamics simulations [27].

Using three different gels with pore sizes $\xi = 5, 6, 7\sigma$, we now test how the mean translocation time is modified by the gel. As expected, Fig. 2 demonstrates that the presence of the gel increases the mean translocation time for all polymer stiffnesses used, with denser gels having a larger impact. The inset shows the same data where the translocation times are rescaled by the mean translocation time $\langle\tau_{\rm ng}\rangle$ in free solution (no gel) conditions, so that

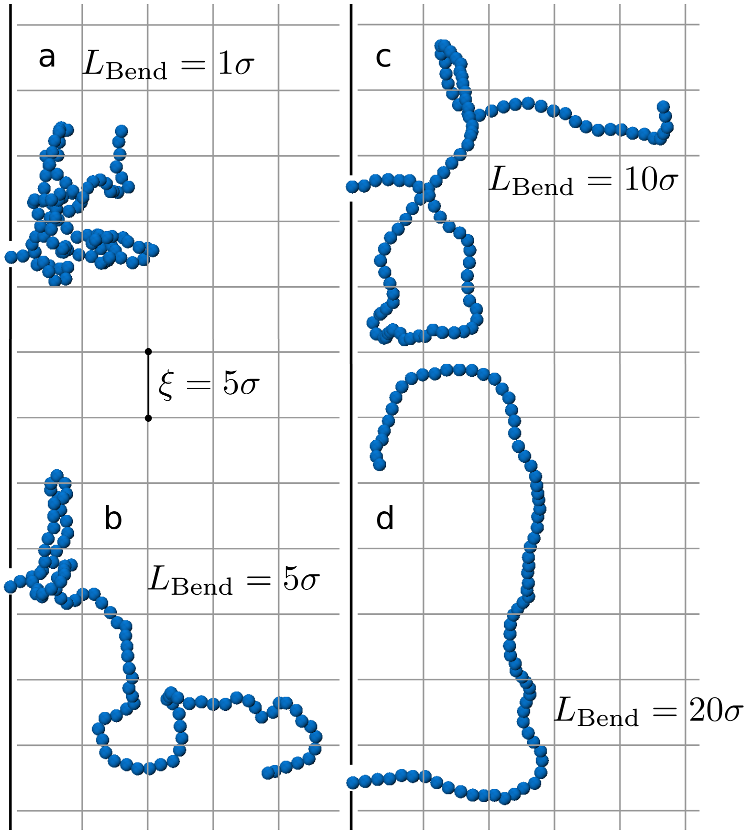


Figure 3: (Color Online) Simulation snapshots for $L_{\text{Bend}} = 1, 5, 10$ and 20σ (a,b,d,c respectively) for the $\xi = 5\sigma$ gel. These are taken at the last simulation frame when $s = N = 100$.

only the combined gel-stiffness effects remain, while the x-axis now shows the ratio L_{Bend}/ξ between the bending length and the pore size. Surprisingly, the data is non-monotonic suggesting a change in physical mechanisms when we move from flexible to rigid molecules. We also note that the maximum is found around $L_{\text{Bend}}/\xi = 2 - 3$ for all gel pore sizes.

The polymer radius-of-gyration being much larger than the pore size of the gel, the reptation model should normally be used to describe its motion inside the gel [28]. In a typical reptation picture, the gel pores are filled in a sequential manner which defines the primitive path—or reptation tube. The polymer moves inside this fictitious tube of constraints imposed by the gel. The formation of hernias [29] (essentially lateral tube leakages) occur at a high entropic cost and can normally be neglected at or near equilibrium. Thus, the reptation tube is evolving by having the leading polymer end explore connected gel pores.

However contrary to typical gel electrophoresis, here the polymer is pushed from behind and the leading end of the polymer is not directly affected by the field. Consider first the flexible limit ($L_{\text{Bend}} \to 0$). The gel pore adjacent to the nanopore is the first to receive translocated monomer beads, but it will eventually be filled. If the polymer is flexible, the leading end cannot feel the effect of the driving force and thus the exploration of different gel pores is mostly a thermal process accelerated by monomer crowding. In this translocation picture, if the incoming monomers arrive at a higher rate than the rate of formation of the reptation tube, some gel pores may become overcrowded and *trans* monomers may be pushed through other openings to accommodate translocation—leading to the formation of hernias. Figures 3 a) and b) show typical conformations at the last translocation step for relatively flexible polymers. We clearly observe an accumulation of monomers in the first few gel pores near the nanopore exit, and the monomers overflow into neighbouring gel pores via hernias. The $L_{\text{Bend}} = 5\sigma$ case produces fewer hernias than when $L_{\text{Bend}} = 1\sigma$, as expected.

In the rodlike limit ($L_{\text{Bend}} \to \infty$), on the other hand, the polymer moves as a solid stick which comes into steric contact with the gel fibres. The growth of the primitive path is no longer a diffusive process, but is rather driven by the applied force in the nanopore. The effect of the gel on the *trans*-side is to supply "gel resistance" contact points to the polymer as it slides across the fibres. Obviously, gels with smaller pores have a higher density of resistance sites (roughly one fibre per ξ of polymer backbone length on the *trans* side). This is very similar to gel electrophoresis of DNA fragments in tight gels [28, 30]

For the $L_{\text{Bend}} = 10$ and 20σ cases shown in Figs. 3 c) and d), hernias are almost non existent (what appears to be a hernia for $L_{\text{Bend}} = 10\sigma$ is actually a polymer loop across multiple pores in the direction perpendicular to the page). The polymer is pushed in the gel and follows a path of least resistance, which entails that sharp 90° bends do not generally have time to occur before the leading end finds the opening opposite the entrance (straight passage). Although such bends do sporadically occur (as shown in the snapshots), their rarity creates what is an effective increase in the persistence length.

With these two limiting mechanisms in mind, one can attempt to interpret the non-monotonic data of the inset of Fig. 2. Increasing the rigidity from the fully flexible case has

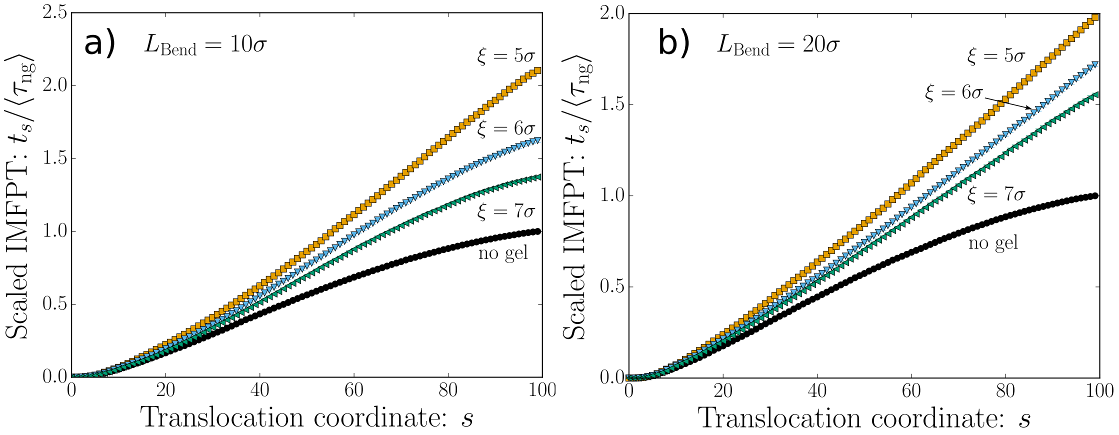


Figure 4: (Color Online) The Incremental Mean First Passage time t_s as a function of the translocation coordinate s. The curves show data for different gel pore sizes ξ. The time t_s is scaled with the mean final ($s = N = 100$) translocation time in the no-gel condition. Data is for polymers with bending lengths a) $L_{\rm Bend} = 10\sigma$, and b) $L_{\rm Bend} = 20\sigma$.

two effects: the gel pores get filled with fewer monomers and the energetic cost to forming hernias increases. The combined effect of these two factors increases the translocation time (compared to the no-gel case). The translocation time increases with stiffness until $L_{\rm Bend}/\xi > 2 - 3$, i.e. until the the polymer rigidity starts limiting the formation of hernias. The governing process then transitions into the rod-like scenario described above. In the latter case, increasing the stiffness actually reduces the density of contact points between the polymer and the gel fibres — thus slightly decreasing gel resistance and speeding up translocation.

We now turn our attention to the dynamics of translocation given by the Incremental Mean First Passage Time (IMFPT) t_s, which is the time that it takes for a given monomer s to reach (or enter) the nanopore *for the first time*. We plot in Fig. 4 the IMFPT for all the monomers of a polymer with a stiffness of a) $L_{\rm Bend} = 10\sigma$ and b) $L_{\rm Bend} = 20\sigma$. Again, we rescaled the time using the no-gel mean translocation time $\langle\tau_{\rm ng}\rangle$. Clearly, the presence of the gel slows down translocation during the entire process (i.e., for all monomers).

The sigmoidal curves for the two no-gel cases in Fig. 4 are typical of the translocation process described by the Tension-Propagation theory [31, 32, 33]. In particular, the instantaneous translocation rate is determined by the number of *cis*-side monomers that contribute to the frictional drag because they are part of the chain section that is under tension. The high translocation rates at the beginning and at the end of the process are interpreted as situations where the monomer being driven (in the nanopore) does not need to pull on many *cis* monomers because the tension front has just begun to grow ($s \approx 0$) or because the process is almost complete ($s \approx N$). The translocation rate (the inverse of the slope) being the lowest at the inflection point (here at $s^* \approx 30$), this point tells us when the tension front has reached the last monomer on the *cis*-side. In other words, this corresponds to the maximum number of *cis* monomers ($= N - s^* \approx 70$ in this example) set in motion due to the applied force in the nanopore. Beyond s^*, the no-gel translocation rate increases because the number of moving monomers on the *cis*-side (hence the friction) is decreasing until the end.

Figure 4 shows that the presence of the gel greatly diminishes the acceleration found at the end of the translocation. In fact, in many cases the translocation rate is essentially constant beyond the initial acceleration phase (i.e., for $s > s^*$) – see for example the $\xi = 5\sigma$ curves in Figs. 4 a and b. It is as if the number of monomers contributing to the drag effectively remains constant from then on: as translocation proceeds, the *cis* monomers transition from being *pulled* to being *pushed* on the *trans*-side. This is consistent with the gel penetration picture of a rod-like polymer since the gel fibres help prevent the semi-flexible segments from buckling. As we will show, this can also be seen as an increase in the effective persistence length L'_{p} due to the gel.

To better understand the IMFPT curves, we can also examine the polymer conformations immediately after translocation is complete, i.e., when the last monomer $s = N = 100$ has exited the nanopore. In Fig. 5, we plot the ensemble mean distance between each of the $N = 100$ monomers and the nanopore for a few semiflexible cases with a gel of pore size $\xi = 5\sigma$ as well as for the no-gel case. For the flexible cases with $L_{\mathrm{Bend}} < 10\sigma$, the post-translocation *trans*-conformations are more extended for the no-gel case than for the case with a gel. This is consistent the fact that hernias form and the polymer conformation collapses when a flexible polymer is pushed inside a gel – see Fig. 3a. In the opposite scenario where $L_{\mathrm{Bend}} \geq 10\sigma$, the polymer is stiff (compared to the gel pore size ξ) and it is

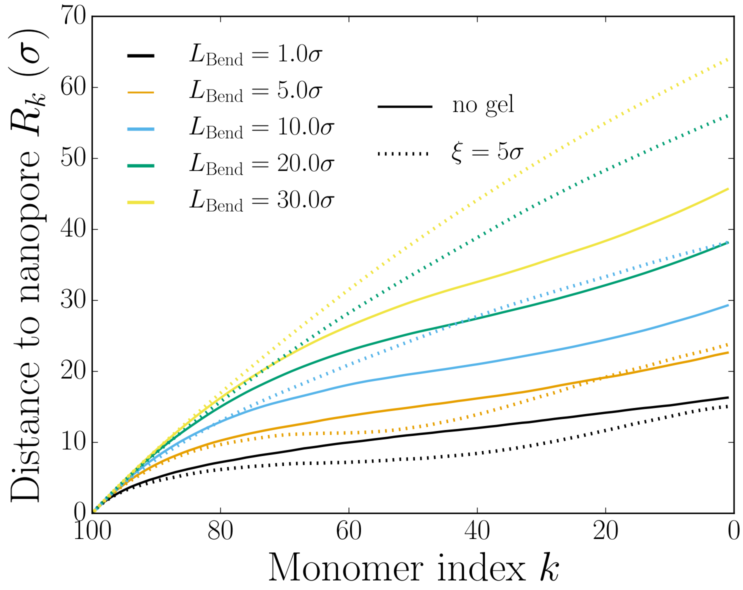


Figure 5: (Color Online) The monomer distance to the nanopore, R_k, as a function of the monomer index k. The conformation data were taken from the last simulation frame of the translocation, i.e. when s=N=100. Note that the last monomer $k = 100$ is inside the nanopore.

more extended when translocating inside a gel than in free solution; see also Figs. 3c-d. In a sense, it is as if the gel fibres act in a way which increases the chain effective persistence length.

As a way to quantify the compression effect for flexible polymers and the alignment effect for the more rigid polymers, we can calculate the effective persistence length from the ensemble of conformations taken at the last simulation frame. The autocorrelation function of the unit bond vectors is calculated and fitted using Eq. 4 in order to obtain the effective persistence length L'_{p}. In Fig. 6, we plot L'_{p} as a function of the bending length L_{Bend} (the former is scaled by its value for the no-gel case, while the latter is scaled by the pore size ξ). The polymers inserted into a gel exhibit a net increase in their effective persistence length when the polymer bending length L_{Bend} is larger than about 1.5 ξ. For the more flexible polymers, however, we observe a (small) decrease in the persistence length, which is consistent with the chain collapse mechanism discussed previously and observed in Fig. 3a. Remarkably, the rescaled data essentially provides a universal curve in Fig. 6,

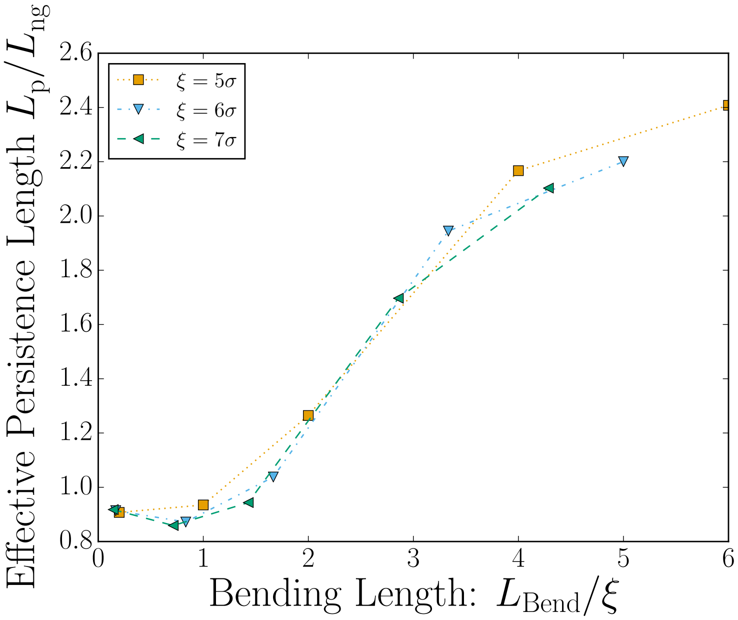


Figure 6: (Color Online) The fitted effective persistence length L'_{p}, scaled by the value found in absence of a gel, $L'_{\mathrm{p_{ng}}}$ plotted as a function of the bending length L_{Bend} scaled by the gel pore size ξ. The conformation data were taken from the last simulation frame of the translocation, i.e. when $s = N = 100$.

clearly demonstrating that the fundamental control parameter is the ratio L_{Bend}/ξ in this problem.

4 Conclusion

We have presented a study of the physical mechanisms at play when a polymer driven solely by a force localized inside a nanopore enters a region containing a gel matrix. Our simulation data have shown that although gels increase the translocation times, there are in fact two distinct regimes depending on the ratio of the polymer bending length L_{Bend} and the gel pore size ξ: these two regimes can called the *flexible regime* when the polymer bending length is smaller than the gel pore size, and the *rod-like regime* when the bending length is larger than the gel pore size.

By considering the mean translocation time $\langle\tau\rangle$ in the presence of a gel scaled by the control value $\langle\tau_{\mathrm{ng}}\rangle$ in absence of a gel, we found a nonmonotonic dependence on the polymer rigidity. In the flexible regime the scaled translocation time $\langle\tau\rangle/\langle\tau_{\mathrm{ng}}\rangle$ *increases* with increasing polymer rigidity, whereas it actually *decreases* with rigidity in the rod-like regime. This entails the existence of a maximum, which occurs when the bending length L_{Bend} is about twice the gel pore size ξ in our simulations.

We investigated the polymer conformations at the instant when the translocation process is complete. We analyzed the effects of the gel on the final polymer conformations by comparing the statistical effective persistence length L'_{p} to the values obtained in absence of the gel, $L'_{\mathrm{p_{ng}}}$. We obtained a data collapse unto a universal curve when these are plotted as a function of the ratio between the bending length and the gel pore size, thus confirming that this ratio is the key physical parameter. In the flexible regime, the effective persistence length in the presence of a gel is *smaller* than without a gel. We find the opposite effect, that is an *increase* in the effective persistence length in the rod-like regime. We again find a crossover in behaviour which occurs near $L_{\mathrm{Bend}}/\xi = 2$.

These data, together with the visualization of the process using pictures of conformations and animations of the process, support two different physical pictures put forth to describe the process. When the gel pore size is smaller than the statistical bending length, the polymer section entering the gel region "crumble" —or buckle—upon entering the gel pores near the nanopore. To accommodate further progression of translocation, monomers exit into neighbouring pores by the formation of one or more hernias; this results in a *decrease* of the effective persistence length. In the rod-like regime, on the other hand, hernias come with a high energetic cost and the polymer segments driven in the gel have a tendency to remain aligned because they do not have time to explore all of the possible directions. This leads to an increase in the statistical persistence length.

This effective increase in the polymer persistence length leads to mean translocation rate which can be remarkably constant throughout the majority of the process. Our results thus suggest that casting a gel, or a nano-fabricated gel-like structure, where the pore-spacing is smaller than the bending length of the polymer, could be used in applications where a constant translocation rate is sought after, such as polymer sizing.

5 Acknowledgements

The Langevin Dynamics simulations were performed with the ESPResSo package [26] on SHARCNET (www.sharcnet.ca). Simulation visualization and plotting were done using VMD and Matplotlib [34, 35]. D.S. is supported by the NSERC-CREATE training program in Quantitative Biomedicine. This work is supported in part by an NSERC grant RGPIN/046434-2013 to G.W.S.
The authors have declared no conflict of interest.

References

[1] Muthukumar, M., *Polymer Translocation*, CRC Press 2011.

[2] Maitra, R. D., Kim, J., Dunbar, W. B., *Electrophoresis* 2012, *33*, 3418–3428.

[3] Fyta, M., *Journal of Physics: Condensed Matter* 2015, *27*, 273101.

[4] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., *Molecular Biology of the Cell*, Garland Science, New York, 4 ed. 2002.

[5] Gopinathan, A., Kim, Y. W., *Phys. Rev. Lett.* 2007, *99*, 228106.

[6] Cao, W.-P., Sun, L.-Z., Wang, C., Luo, M.-B., *J. Chem. Phys.* 2011, *135*, 174901.

[7] de Haan, H. W., Slater, G. W., *Phys. Rev. E* 2014, *90*, 020601.

[8] Squires, A. H., Hersey, J. S., Grinstaff, M. W., Meller, A., *J. Am. Chem. Soc.* 2013, *135*, 16304–16307.

[9] Tang, Z., Liang, Z., Lu, B., Li, J., Hu, R., Zhao, Q., Yu, D., *Nanoscale* 2015, *7*, 13207–13214.

[10] Waugh, M., Carlsen, A., Sean, D., Slater, G. W., Briggs, K., Kwok, H., Tabard-Cossa, V., *Electrophoresis* 2015, *36*, 1759–1767.

[11] Dubbeldam, J., Rostiashvili, V., Milchev, A., Vilgis, T., *Phys. Rev. E* 2012, *85*, 041801.

[12] Lehtola, V. V., Linna, R. P., Kaski, K., *Phys. Rev. E* 2008, *78*, 061803.

[13] Lehtola, V., Kaski, K., Linna, R., *Phys. Rev. E* 2010, *82*, 031908.

[14] Linna, R., Kaski, K., *Phys. Rev. E* 2012, *85*, 041910.

[15] Luo, K., Ala-Nissila, T., Ying, S.-C., Metzler, R., *EPL* 2009, *88*, 68006.

[16] Dubbeldam, J. L. A., Rostiashvili, V. G., Milchev, A., Vilgis, T. A., *Phys. Rev. E* 2013, *87*, 032147.

[17] Bhattacharya, A., Morrison, W. H., Luo, K., Ala-Nissila, T., Ying, S.-C., Milchev, A., Binder, K., *Eur. Phys. J. E* 2009, *29*, 423–429.

[18] Lehtola, V., Linna, R., Kaski, K., *Phys. Rev. E* 2010, *81*, 031803.

[19] Lehtola, V., Linna, R., Kaski, K., *EPL* 2009, *85*, 58006.

[20] Luo, K., Ala-Nissila, T., Ying, S.-C., Bhattacharya, A., *Phys. Rev. Lett.* 2007, *99*, 148102.

[21] Huopaniemi, I., Luo, K., Ala-Nissila, T., Ying, S.-C., *J. Chem. Phys.* 2006, *125*, 124901.

[22] de Haan, H. W., Sean, D., Slater, G. W., *Phys. Rev. E* 2015, *91*, 022601.

[23] Slater, G., Holm, C., Chubynsky, M., De Haan, H., Dubé, A., Grass, K., Hickey, O., Kingsburry, C., Sean, D., Shendruk, T., Z., L., *Electrophoresis* 2009, *30*, 792–818.

[24] Weeks, J. D., Chandler, D., Andersen, H. C., *J. Chem. Phys.* 1971, *54*, 5237–5247.

[25] Grest, G., Kremer, K., *Phys. Rev. A* 1986, *33*, 3628–3631.

[26] Limbach, H.-J., Arnold, A., Mann, B. A., Holm, C., *Comput. Phys. Commun.* 2006, *174*, 704–727.

[27] Adhikari, R., Bhattacharya, A., *J. Chem. Phys.* 2013, *138*, 204909.

[28] Viovy, J.-L., *Rev. Mod. Phys.* 2000, *72*, 813–872.

[29] Sean, D., Slater, G. W., *Electrophoresis* 2010, *31*, 3446–3449.

[30] Duke, T., Semenov, A., Viovy, J., *Phys. Rev. Lett.* 1992, *69*, 3260.

[31] Sakaue, T., *Phy.s Rev. E* 2007, *76*, 021803.

[32] Sakaue, T., *Phys. Rev. E* 2010, *81*, 041808.

[33] Saito, T., Sakaue, T., *Eur. Phys. J. E* 2011, *34*, 1–8.

[34] Humphrey, W., Dalke, A., Schulten, K., *J. Molec. Graphics* 1996, *14*, 33–38.

[35] Hunter, J. D., *Comput. Sci. Eng* 2007, *9*, 90–95.

8

Driven polymer translocation into a crosslinked gel: Tension-Propagation theory

David Sean, Gary W. Slater

This is a short theoretical addendum to the manuscript found in the previous chapter. A high-field Tension-Propagation model is derived to include the effects of polymer stiffness. Some of the gel effects on the *trans*-side can be explicitly assessed using this model.

Some of the contents of this chapter will be submitted for peer review; the future manuscript will offer a more complete derivation of the Tension-Propagation theory of semiflexible polymers.

Driven polymer translocation into a crosslinked gel: Tension-Propagation theory

David Sean and Gary W. Slater
Physics Department, University of Ottawa,
Ottawa, Ontario, Canada, K1N 6N5

(Dated: January 4, 2017)

INTRODUCTION

In the previous Chapter, we investigated the physical mechanisms which explain how a semiflexible polymer enters a gel when driven via a nanopore. The purpose of this short chapter is to supplement this investigation with the Tension-Propagation (TP) theory. Due to the fact that the role of polymer rigidity was explored in Chapter 7, we now derive a general TP model for semiflexible polymers. This is done via the addition of a new parameter, N_t, which is the number of monomers translocated to the *trans*-side of the membrane which are being pushed by the monomer being driven in the pore. Comparing the predictions of our TP-theory to the results of Langevin Dynamics simulations allows us to estimate how many *trans* monomers need to be taken into account. Furthermore, we can better understand the effect of a *trans*-side crosslinked gel by observing how this parameter is affected.

In this exploration, we employ a generic polymer model but we do not try to match the translocation Péclet number with that in experimental work. This is due to the fact that the introduction of a bending length scale will directly affect the coarse-graining procedure. Since it is not our aim to modify the granularity of our model for every polymer of varying stiffness, and since the stiffness of DNA is a material property with limited experimental control, we shall refrain from direct comparison with DNA. This study aims to map-out the effect of semiflexibility unto a generic polymer model.

As with the previous chapter, we use $k_{\mathrm{B}}T = 1\epsilon$ and $F\sigma/k_{\mathrm{B}}T = 1$ here. One caveat to keep in mind is the applicability of our TP theory which is normally limited to the *highly-driven* limit, $F\sigma/k_{\mathrm{B}}T \gg 1$.

TENSION-PROPAGATION

Theory

We propose to incorporate the effect of polymer stiffness into a Tension-Propagation (TP) model [1–4] via a variable N_t defined as the number of monomers which need to be pushed on the *trans*-side during translocation. This is in contrast to the TP theory for a flexible polymer model which neglects the effect of the monomers that have translocated. As usual, the number of monomers being dragged on the *cis*-side is determined from the length of the

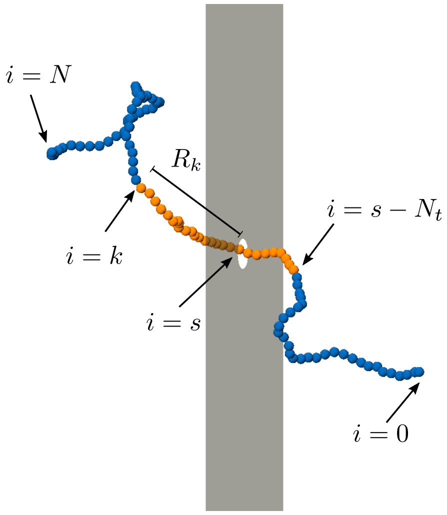


FIG. 1. A schematic explaining the variables used in this chapter. A $N = 100$ semiflexible polymer is driven via force F applied to the monomer s inside the pore. This tension travels down to monomer k creating a tension segment on the *cis*-side of length R_k that get *pulled*. Due to polymer rigidity there are N_t monomers on the *trans*-side that are being *pushed* by the force on monomer s.

segment under tension. In the highly driven regime this forms a *stem* of length

$$R_k = (k - s)\sigma, \tag{1}$$

where k is the monomer index at the tension front and s is the monomer index in the pore (see Fig. 1). We will use the dimensionless variable $\tilde{R}_k = R_k/\sigma$ for simplicity.

Neglecting friction effects from the nanopore and in an overdamped regime, the driving force acting on monomer s inside the pore is balanced by the $(k - s)$ monomers being *pulled* on the *cis*-side in addition to the N_t monomers being *pushed* on the *trans*-side. This gives a translocation rate of

$$\frac{ds}{dt} = \frac{F}{\zeta\sigma} \left(k - s + N_t\right)^{-1} \tag{2}$$

where the applied force F, the monomer friction coefficient ζ and the monomer size σ gives the natural time units $\tau_0 = \zeta\sigma/F$. Figure 1 shows the $(k - s + N_t)$ moving monomers coloured in orange. This translocation rate differs from what is found in flexible polymer models because of the new variable N_t which we add here to account for monomer correlations that *persist* over a length scale longer than the monomer size in a semiflexible chain.

In order to find N_t, we can imagine a stiff rod-like polymer being driven from its middle monomer. Since the polymer moves as a single unit, the middle monomer would need to pull $N/2$ monomers on one side while pushing $N/2$ monomers on the other. As the polymer stiffness is decreased, the number of monomers being pulled would remain unchanged, but the *pushing capacity* for the middle monomer would certainly decrease. For a sufficiently flexible polymer, the front half would buckle and bend out of the way to allow for further translocation, as observed in ref [5]. Building on the idea of a polymer's pushing capacity, we use $N_{\max}$ as the maximum number of monomers which can be pushed by a force applied downstream along the polymer backbone. A first approximation for this property would be the persistence length L_{p}, but we will demonstrate that although L_{p} is a relevant length scale, this is not generally the case. In this manuscript, we will use $N_{\max}$ as a fitting parameter.

The number of monomers being pushed is at first limited by the amount present on the *trans*-side, s. Thus, we will assume that the number of monomers being pushed grows with the translocation coordinate s and later plateaus at its maximum capacity $N_{\max}$:

$$N_t = \begin{cases} s & \text{for } s < N_{\max} \\ N_{\max} & \text{for } s \geq N_{\max} \end{cases}. \tag{3}$$

By taking the time derivative of Eq. 1, we can write the rate of growth of the tension front k as

$$\frac{dk}{dt} = \left(1 - \frac{d\tilde{R}_k}{dk}\right)^{-1} \frac{ds}{dt}. \tag{4}$$

If we assume the process is highly driven such that the monomers do not significantly move from their initial location before being pulled by the applied force, we can obtain an additional expression relating the distance R_k between monomer k and the nanopore directly from the initial polymer conformations. These conformations are obtained via simulations of a discrete bead-spring polymer with the first monomer fixed inside the nanopore. This allows us to numerically obtain $d\tilde{R}_k/dk$ by centered differences.

Using the initial condition that $s = 0$ and $k = 0.5$ (we use $k = 0.5$ to avoid any mathematical divergence), Eqs. 2 and 4 can be integrated numerically to obtain their respective time evolutions $s(t)$ and $k(t)$ for a given set of initial conformations. A time step $\Delta t = 0.01\tau_0$

produces stable numerical results sufficiently similar to the analytical expression which can be obtained when $N_t = 0$, as described in the Introduction (Chapter 1).

Tension-Propagation in absence of a gel

We first examine the behaviour of the aforementioned TP theory that includes semi-flexibility. In order to use Eq. 4, we need to evaluate $d\tilde{R}_k/dk$. Instead of using an analytical expression, we will directly make use of the same initial conformations of the LD simulations. Figure 2 shows the results for random walk polymer conformations generated by MC simulations. The MC simulations generate self-avoiding walks by having the first monomer tethered on a flat wall. If additional monomers are found to cross the wall, or are found to lie a distance smaller than a bead size $\Delta r < \sigma$ to other existing monomers, the walk is restarted. Semiflexibility is included by weighting monomer positions by an associated Boltzmann factor $e^{-U_{\rm Bend}/k_{\rm B}T}$ using a Metropolis rejection scheme. The bending potential has a harmonic form

$$U_{\rm bend} = \frac{1}{2}\kappa\left(\phi - \pi\right)^2, \tag{5}$$

where κ is the bending constant. In three dimensions, this form gives an approximate persistence length of $L_{\rm p}/\sigma \approx \kappa/k_{\rm B}T$ [6]. These same bending potential and conformations are used to initialize the LD translocations. From these curves, we can numerically evaluate the slopes $d\tilde{R}_k/dk$ and interpolate the values for an arbitrary (i.e., non-integer values of) k.

This enables us to directly test the predictions of the TP theory and qualitatively determine the effect of having $N_{\rm max} \neq 0$. In Fig. 3a, we use the conformations generated for a bending parameter $\kappa = 5.0k_{\rm B}T$ for different values of $N_{\rm max}$. As expected, the translocation time increases with $N_{\rm max}$. This is simply due to a higher number of monomers (coming from the *trans*-side) that contribute to the drag and slow down the translocation process. Furthermore, the curvature near the end of the process is affected by $N_{\rm max}$; the much reduced terminal rate $(dt/ds)^{-1}$ reflects the fact that the depletion of *cis*-monomers does not decrease the amount of drag so much because the N_t *trans* monomers given by Eq. 3 participate in the total drag. The finite instantaneous final translocation rate can be evaluated using Eqs. 2 and 3 with $k = s = N$:

$$\left.\frac{ds}{dt}\right|_{s=N} = \frac{1}{\tau_0 N_{\rm max}}, \tag{6}$$

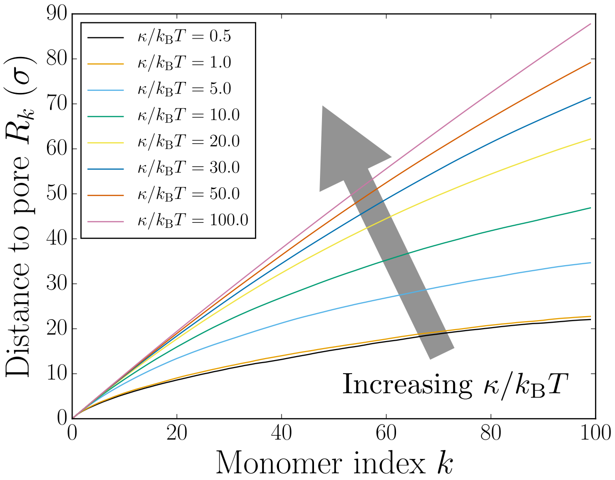


FIG. 2. The mean distance to the pore R_k as a function of the monomer index k as obtained from 1000 half-space self-avoiding walks of $N = 100$ steps generated using a Monte-Carlo approach. These are used as initial conformations for the LD simulations as well as to determine $d\tilde{R}_k/dk$ for the TP theory.

a direct consequence of *trans*-monomers being pushed.

Figure 3b focuses solely on the impact of the initial conformations. The parameter $N_{\mathrm{max}} = 0$ is left at zero while different initial conformations corresponding to varying degrees of semiflexibility are used. Note the divergence of the final translocation rate (the slope is zero at $s = N$) which is expected when only *cis*-monomers contribute to the drag. Interestingly, increasing κ results in a net decrease in the Tension-Propagation time. Since the introduction of semi-flexibility removes small scale fluctuations along the backbone, the tension front travels faster down the chain. The markers in Fig. 3a-c show the point in time when the tension front has traveled down the entire chain. For a flexible polymer, the post-propagation time can be taken as a small correction to the total translocation time [1] whereas the opposite can be observed for a sufficiently stiff polymer (comparing $\kappa = 50k_{\mathrm{B}}T$ to $\kappa = 1k_{\mathrm{B}}T$). This is similar to the effect found when a polymer is initially contained in a confining tube [7].

Figure 3c shows TP predictions for different semi-flexible polymers where we make the assumption that $N_{\mathrm{max}} = L_{\mathrm{p}}/\sigma = \kappa/k_{\mathrm{B}}T$. This situation shares the combined features of the two preceding situations shown in Figs. 3a and b.

Thus, we model the effect of semiflexibility via two distinct effects: i) the initial confor-

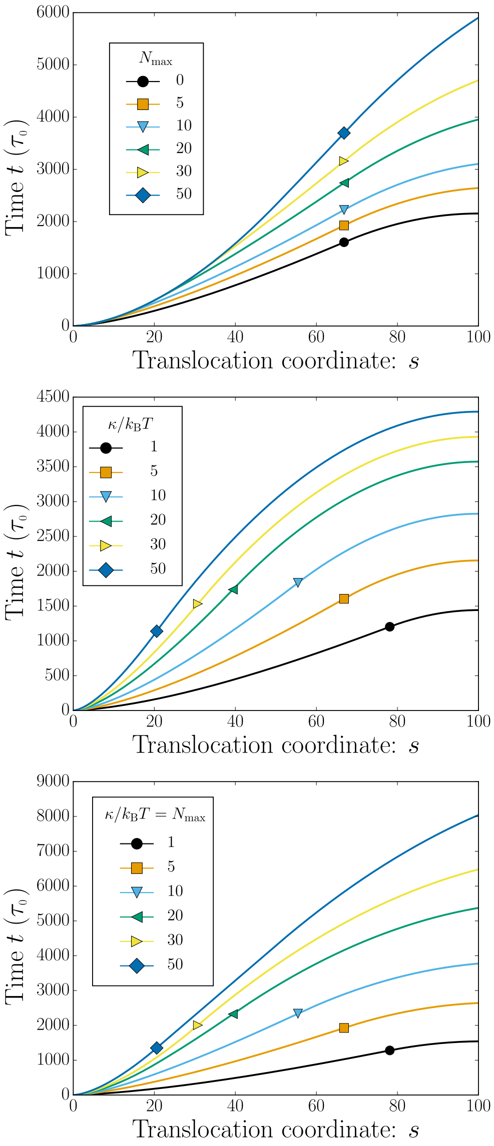


FIG. 3. (Color Online) The time t needed to reach a translocation coordinate s as predicted by TP theory. There are up to $N_{\max}$ monomers on the *trans*-side that contribute to the total drag. The marker shows where the tension has propagated to the end of the chain. a) We use initial conformations generated using $\kappa = 5k_{\mathrm{B}}T$ and vary $N_{\max}$. b) The effect on the *trans*-side is neglected ($N_{\max} = 0$) to only show the influence of initial conformations with different values of κ. c) Both effects are combined using $N_{\max} = L_{\mathrm{p}}/\sigma = \kappa/k_{\mathrm{B}}T$. Note the different time axes.

mations and ii) an additional parameter to account for the number of monomers that must be pushed on the *trans*-side. The former affects the rate of tension-propagation, whereas the latter affects (among other things) the final translocation rate (at $s = N$).

SIMULATIONS

The simulation model is identical to that used in Chapter 7. We use the time scale $\tau_{\mathrm{sim}} = \tau_0 = \zeta\sigma/F$ as these units naturally arise from the Tension-Propagation theory.

Translocation to free solution

Using $N_{\mathrm{max}} = L_{\mathrm{p}}/\sigma = \kappa/k_{\mathrm{B}}T$ as a first approximation gave us a general idea of how the translocation profile is affected by semiflexibility effects in the TP theory (see Fig. 3c). This can be directly verified by turning to LD simulations as a means of generating complete profiles of the translocation dynamics.

In Fig. 4a we plot the incremental mean first passage time (IMFPT) [8] as a function of the translocation coordinate s. From these simulation data, we can fit the TP predictions using N_{max} as a free parameter. The best fits are also shown Fig. 4a.

Due to the fact that the TP theory uses the same $\{R_k\}$ values as were used to initialize the simulations, the effect of N_{max} will be to decrease the translocation rate as a consequence of including the *trans*-monomers into the total drag (see Fig. 3a). Although the translocation rate is indeed decreased over the process, the end of the process with a rate given by Eq. 6 stands out. Thus we can attempt to extract from the fits the influence of the *trans* monomers on the translocation dynamics.

Close inspection of Fig. 4a reveals that the fitted lines tend to slightly underestimate the final slope (at $s = N = 100$). The fitted value of N_{max} cannot simply be increased to match this slope due to the overestimates at earlier s (as an example, see the $\kappa/k_{\mathrm{B}}T = 50$ data near $s \approx 40$). When the curves are normalized by the final translocation time (data not shown) the fitted theory overshoots (at low κ) and undershoots (at high κ) the simulation data. This is consistent with a compromise between two physical phenomena, which is to be somewhat expected since the simulations were not quite carried out in the highly driven regime.

Various values of $\kappa/k_{\mathrm{B}}T = (0.5 - 100)$ are fitted and the results for N_{max} are plotted in Fig. 4b. The plot shows that the previous assumption that $N_{\mathrm{max}} \sim \kappa/k_{\mathrm{B}}T$ does not hold. On the one hand, we know that the fit is bound by $N_{\mathrm{max}} \leq N$ and we do expect an eventual plateau to occur. However, the results are sub-linear much sooner than I expected,

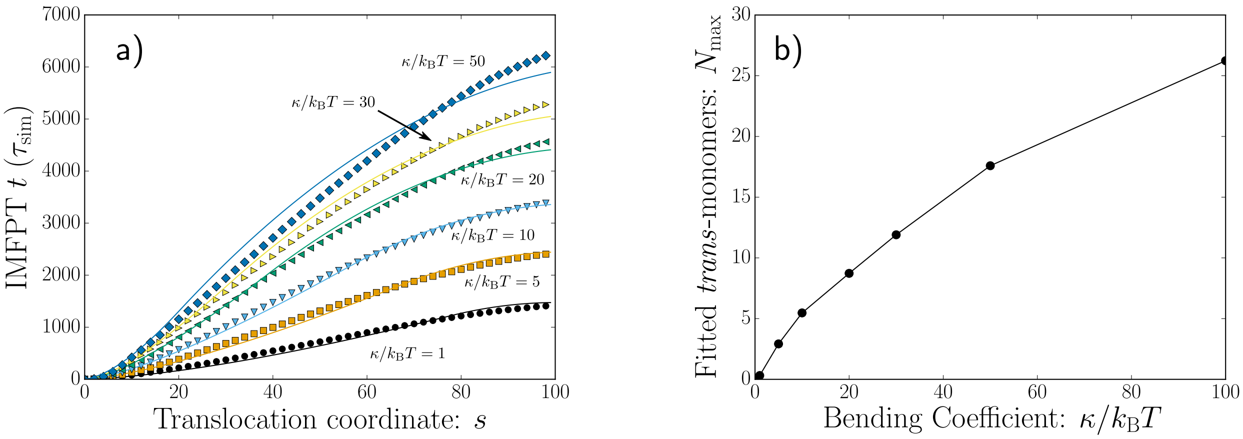


FIG. 4. a) Incremental mean first passage time to reach a tranlsocation coordinate s for semiflexible polymers simulated at varying values of κ (symbols). Using the same initial conformations, the TP theory is fitted using N_{max} as a free parameter. b) The N_{max} fitted values are plotted as a function of $\kappa/k_{\mathrm{B}}T$.

which may again come from effects arising from non-zero thermal fluctuations. The global behaviour is nonetheless encouraging, the growth of N_{max} with κ indicates that with the initial conformations already accounted for, the decrease of the translocation rate can be captured by accounting for *trans* monomers.

Cross-linked gel on the *trans* side

When a gel is placed on the *trans*-side, we expect the fibres to increase the effective persistence length of the polymer (as was reported in Chapter 7) and potentially yield an effective increase in the number of moving *trans* monomers. Figure 5a shows LD simulation results and a fitted estimate of N_{max} using TP theory for a semiflexible polymer with $\kappa/k_{\mathrm{B}}T = 10$. From the fits, we can clearly conclude that the number of monomers needing to be pushed is substantially increased by the presence of the gel.

For the $\kappa/k_{\mathrm{B}}T = 10$ case plotted in Fig. 5a, the value of N_{max} is increased approximately four-fold by including the very porous $\xi = 7\sigma$ gel. Recall that in the last chapter the persistence length was found to be increased by *maximum* of about ×2, as measured from the monomer conformations at the end of the process. The analysis presented here suggests that dynamical correlations (rather than static) are quite magnified by the gel. The increase

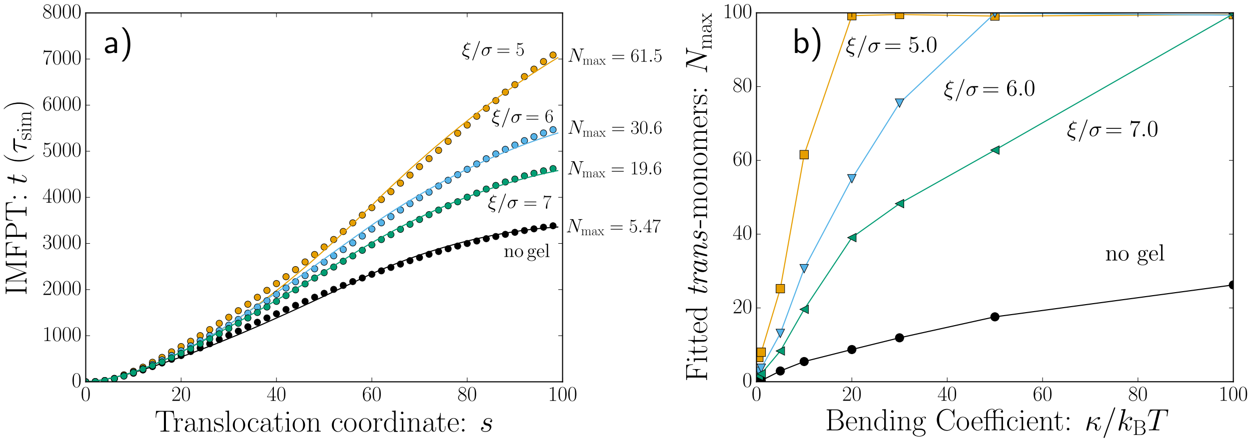


FIG. 5. Simulations and theory with a gel on the *trans*-side. a) The IMFPT plots for $\kappa = 10k_{\mathrm{B}}T$. Simulation data are shown with symbols and the theory fitted with $N_{\max}$ as a free parameter is shown as solid lines; b) The fitted values of $N_{\max}$ as a function of $\kappa/k_{\mathrm{B}}T$.

in $N_{\max}$ is also amplified with increasing bending constant κ. Figure 5b shows that including a gel strongly changes how κ affects $N_{\max}$. For example, the gel with $\xi = 5\sigma$ causes an $N_{\max}$ increase of about ×11 when the polymer has a bending coefficient of $\kappa/k_{\mathrm{B}}T = 10$.

The saturation of $N_{\max}$ limits the capacity of the theoretical fits to capture supplemental increases of the translocation time (see Fig. 5b). If $N_{\max}$ were unbounded by the range $(0, N)$, it would for some cases (small gel pore size ξ with high polymer rigidity κ) take on a value $N_{\max} > N$, which would be unphysical.

This provides direct evidence that frictional contributions, which are currently not accounted for, should be investigated. The TP model is not designed to capture frictional contributions from the gel and is thus rather limited in this situation. In Chapter 7, we found evidence of a regime change: i) the gel increases the translocation time via increased *crowding* in one regime, and ii) the gel increases the translocation time via increased *friction* along the backbone in another.

This is consistent with the picture described here: the TP model breaks down when frictional effects dominate over crowding effects. More work is thus needed in order to ascertain how the gel fibres and pore friction affect the fitted values of $N_{\max}$.

CONCLUSION

A Tension-Propagation model is derived to account for semiflexible polymers. We assume that polymer rigidity changes the manner in which the tension propagates down the chain but only implicitly via the static polymer properties (the initial conformations $\{R_k\}$). For the moment, the initial conformations are not fully integrated into the model, but must rather be generated by an external means, for example, via Langevin Dynamics or Monte-Carlo simulation data.

In addition to affecting initial conformations, polymer rigidity introduces correlations between monomers along the chain. We model these correlations as a non-zero contribution to drag from monomers on the *trans*-side. We use the latter as a fitting parameter and compare the TP theory to the dynamics extracted from LD simulations.

This is investigated for polymers that translocate to free-space and when there is a crosslinked gel on the *trans*-side. When translocations are performed without a gel, the agreement between the LD simulations and the TP theory is rather remarkable considering the fact that the theory is derived for the highly driven regime. When flexible polymers were used in previous comparisons between the TP theory and the simulation data, quantitative agreement was only found in the limit $F_{\mathrm{d}}\sigma/k_{\mathrm{B}}T \gg 1$. The data presented here have a ratio of $F_{\mathrm{d}}/k_{\mathrm{B}}T = 1.0$ which usually fails to yield even qualitative agreement when $N = 100$.

We hypothesize that the introduction of semiflexibility reduces small scale fluctuations and that this helps increase the range of the force regimes for which the stem approximation is valid (Eq. 1). The local monomer density present in the segment under tension is shown for various situations in Fig. 6. This hypothesis remains to be investigated in future work.

In previous a study we investigated the effect of thermal fluctuations on translocation times and found non-trivial effects. In some cases an increase in frictional pore interactions upon an increase in thermal fluctuations were observed. A TP model with an additional pore friction term η_p was fitted to LD data, and we found that increasing the temperature will increase the value of η_p [7].

For the work presented here, it is quite possible that these pore friction effects are effectively absorbed by the fit via the parameter $N_{\max}$. For example, the fitted value for the $\kappa/k_{\mathrm{B}}T = 50$ curve in Fig. 4a seems to contain a compromise between capturing two separate effects. We must therefore be careful when attributing weight to the physical interpretation

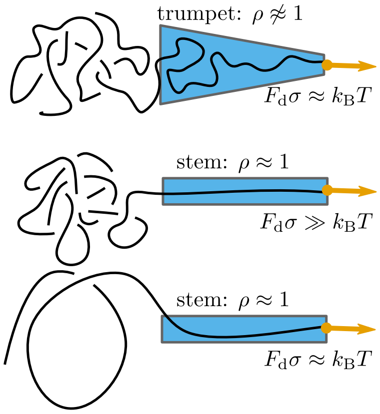


FIG. 6. Polymer segments under tension are highlighted with boxes. Polymer rigidity reduces small scale fluctuations such that the monomer density given by ρ remains well approximated by a stem $\rho \approx 1/\sigma$ for the conditions under study.

of N_{max} until these effects have been accounted for. This is especially true when outside the scope of the theoretical model with the relatively weak driving force $F\sigma/k_{\mathrm{B}}T = 1$.

Nevertheless, since the model captures the dynamics fairly well for translocation into free-space, we also fit simulation data for cases with a crosslinked gel on the *trans*-side. Since N_{max} is the (maximum) number of monomers that can be pushed on the *trans*-side, the fits can be used to verify if the gel increases the number of *trans*-monomers that need to be pushed as translocation progresses. Our data is consistent with this picture: we find that N_{max} increases considerably when a gel is introduced. Interestingly, the increase in N_{max} is much more pronounced than the increase in the effective persistence length L_{p}.

In Chapter 7, the static conformations used in determining L_{p} contained signatures of the underlying gel lattice structure (sharp bends and long straight runs). Static measurements of this data may neglect the dynamical correlations that are involved. The analysis presented here may capture a type of *dynamical* correlation which cannot be extracted from an ensemble of static conformations.

More work is needed to prevent the parameter N_{max} from capturing frictional contributions from both the nanopore and the gel fibres. It is crucial that we decouple the two effects.

Time autocorrelations can be employed to ascertain the true amount by which N_{max} is increased from the gel. Dynamical correlations between the monomer in the pore $i = s$ and a *trans* monomer $j = s - N_{\mathrm{max}}$ may be in fact quite beneficial. Consider the situation where a gel is cast between two closely spaced nanopores (in series) where a DNA is driven such that it is simultaneously present in both pores [9]. A high correlation of DNA movement between the two nanopores can increase the correlations between the signals for valuable error checking in sequencing applications.

---

[1] T. Saito and T. Sakaue, Phys Rev E **85**, 061803 (2012).

[2] T. Saito and T. Sakaue, Eur Phys J E **34**, 1 (2011).

[3] T. Sakaue, Phys Rev E **76**, 021803 (2007).

[4] T. Sakaue, Phys Rev E **81**, 041808 (2010).

[5] A. McMullen, H. de Haan, J. Tang, and D. Stein, under review (2016).

[6] H. W. de Haan and T. N. Shendruk, ACS Macro Letters **4**, 632 (2015).

[7] D. Sean, H. W. de Haan, and G. W. Slater, Electrophoresis **36**, 682 (2015).

[8] H. W. de Haan and G. W. Slater, J Chem Phys **134**, 154905 (2011).

[9] M. Magill, C. Falconer, E. Waller, and H. W. de Haan, Phys Rev Lett **In Press** (2016).

9

Conclusion

The driven transport of linear polymers through small pores was investigated by computer simulations and theoretical modelling. This work was motivated by applications related to DNA analysis, such as sequencing and sizing, but could also be applied to a variety of situations where driven linear polymer translocations are involved.

Since experimental realizations of driven DNA translocation contain a surprisingly wide distribution of translocation times, I have explored ways in which the width of these distributions could potentially be reduced. This was done primarily using coarse-grained Langevin Dynamics simulations.

The simulations provided numerical translocation data under various tested geometries and driving conditions. This allowed me to evaluate the possible advantages and disadvantages of the investigated translocation setups. Naturally, the most promising situations need to be inspected more closely—ideally experimentally—in order to validate my findings.

One of these findings is quite fundamental. I posited that initial conformations contributes significantly to the width of the translocation time distributions. This kind of *molecular individualism* had started to be discussed in the field of translocation at the time when I started my PhD. To the best of my knowledge this hypothesis has not been validated experimentally.

I found the effect of the initial conformations to be quite present when my simulations were tuned such that the translocation time τ_{trans} was much shorter than the polymer relaxation time τ_{relax}. The ratio, which we called the translocation Péclet number (after a suggestion made by Tyler Shendruk upon overhearing a discussion between Hendrick de Haan and myself)

$$P_{\text{t}} = \frac{\tau_{\text{relax}}}{\tau_{\text{trans}}}, \tag{31}$$

could be estimated from published experimental data.

When simulations were performed under conditions that yield a simulation value of P_{t} that was similar to our estimates of the experimental P_{t} value, the initial conformations were found to contribute significantly to the total variance in translocation times. This is quite suggestive that the initial conformations could indeed matter in the lab.

Interestingly, the needed Péclet number was found to occur for a simulation parameter space seldom visited by other groups performing similar coarse-grained simulations. In order to obtain an experimentally relevent P_{t}, we needed to increase the magnitude of the energy gained by a monomer by crossing to the *trans-side* relative to the thermal energy by a factor of 50 (most simulations in the literature have this factor near unity)!

Our resulting simulations were found to be in the highly-driven translocation regime—a situation well modelled by an existing Tension-Propagation theory. This theory was originally proposed by Prof. Sakaue and co-workers who, as it turned out, also mentioned the concept of a translocation Péclet number.

The true causes of the surprisingly wide experimental distribution remains unknown. However, our simulations were consistent with Sakaue's theoretical prediction: perhaps the initial conformations are indeed to blame. Armed with a simulation parameter space calibrated by the experimental value of P_{t}, I thus focused on finding ways to reduce the range of initial polymer conformations.

The physics of confined DNA has attracted a lot of attention from theoretical, simulation and experimental groups. Since experimental teams demonstrated the ability to confine DNA in narrow channels, it became obvious to me that translocation could benefit from confinement—especially in light of limiting the range of initial conformations.

I thus focused my efforts in confining the simulation polymers in a narrow tube prior to translocation and successfully found how to decrease the widths in the translocation time distributions. The simplest scenario for confinement that I found was a semi-infinite tube with a varying diameter as a control parameter (Chapter 3). At the time, I was also investigating slowing down translocation using a crosslinked gel on the *cis*-side and felt that the (reptation) tube could have been a good model for the effects due to the presence of the gel. A Tension-Propagation model was derived for these geometries which agreed with my simulations. Not only were the confining geometries successful in decreasing the distribution width, I accidentally stumbled upon a way to also *increase* the mean translocation time.

The tube idea was further explored in a second manuscript (Chapter 4) where I used a closed cylindrical cavity, as opposed to the semi-infinite tube. By changing the cavity aspect ratio I was surprised to find that the minimum in the width of the distribution of translocation times was obtained with an aspect ratio such that the cavity diameter is twice as long as the cavity length. A model based on Tension-Propagation theory for this situation suggested that the ideal cavity geometry would be similar to a hemisphere where the nanopore is placed at the centre of the circular face.

Gary's Postdoctoral Fellow Hendrick de Haan was working on a translocation scenario where the translocating polymer was kept under a stretched state via two forces applied on the polymer ends. I joined the project after we both realized how the pre-stretching constraints could provide a way to control the initial polymer conformations.

The work in Chapter 5 used the stretching force as a control parameter to modulate the initial conformations. We focused on analyzing the translocation time distributions across different values of P_{t} and different amounts of pre-stretching to map how both initial polymer conformations and Brownian fluctuations affect the resulting variance of translocation times. For completeness, a theoretical Tension-Propagation model for the end-pulling translocation of a prestretched polymer was derived in Chapter 6.

Although my primary aim was to reduce the width of the translocation time distributions, I discovered that some situations (confinement in a long, thin tube for instance) resulted in a considerable increase of the translocation time. This turned out to be an advantage since experimental work also suffers from a high translocation rate which is difficult to decrease without affecting the signal quality. With the main objective of decreasing the translocation rate, I tested how placing a cross-linked gel on the *trans*-side would affect the dynamics. This shared idea was explored experimentally with collaborators [28] and the situation turned out to involve the rather interesting physical picture of pushing a flexible polymer into a region dense with obstacles. In Chapter 7, I used simulations in a more generic parameter space to explore the basic physics of such an unique scenario. As with the preceding situations, I derived a theoretical model in Chapter 8 which accounts for the effect of polymer stiffness.

9.1 Outlook

As discussed above, computer simulations together with theoretical Tension-Propagation models were used in various situations of driven polymer translocation. In many cases there were intricate connections between the two which I tinkered with in order to shine light onto the different roles that they play. In the following sections, I go over situations of how the "cross-talk" between the TP theory and simulations came into play and how it ignited interest in different directions.

9.1.1 Deterministic simulations

The tension propagation models considered in this thesis mostly neglected the monomers located on the *trans*-side. In Chapter 3, Langevin Dynamics simulations were performed where the monomers were removed from the simulation as they entered the *trans*-side. Comparing the simulation results with and without the presence of these monomers enabled me to assess the quantitative impact of neglecting them in the TP theory. It has always been generally considered to be a small correction, and this hypothesis is now supported by my simulations.

Simulation movies of the *deterministic* simulations with $k_{\mathrm{B}}T = 0$ showed that the monomers on the *trans*-side simply accumulated into a dense "ball". Deterministic simulations were rather uninteresting and my use of them was only pedagogical, but it became obvious that if $F\sigma/k_{\mathrm{B}}T \to \infty$ (by $k_{\mathrm{B}}T \to 0$) then perhaps crowding should play an important role. One line of thinking would conclude that they must contribute since the *trans*-monomers could not diffuse away at $k_{\mathrm{B}}T = 0$. As it turns out, crowding effects remain negligible in comparison to the drag from *pulling* the *cis* monomers.

The reverse scenario is however quite interesting. What kind of variance via molecular individualism would be present in what was once a compressed *trans*-polymer? My naïve prediction is that this should give a situation *very* similar to the ideal hemispherical cavity that was discussed in Chapter 4—without the need of an actual geometrical contraint. Reversing translocation is not a new concept but, I have never seen it cast as a way of producing the ideal initial conformations as I am suggesting.

9.1.2 Pore friction

With the exception of Chapter 3, I neglected the addition of a polymer-pore frictional force in the force-balance equation while deriving the TP theory. This was mostly due to the fact that friction brings a linear (N^1) contribution to the translocation time, which becomes negligible next to any higher powers at $N \gg 1$. Interestingly, the issue of N-scaling was historically quite controversial since many groups performing experiments, theory, and simulations were getting different results. The controversy for driven translocation is mostly resolved by the acknowledgement of a long-lasting finite-size effect [26]. This is due to the pore friction term which contributes to a high prefactor thus demanding $N > 10^{5-6}$ monomers to fully disappear.

The physical origin of this frictional term in my simulations was linked to the interactions between the corrugate polymer and the edge of the nanopore. I have attempted to log the forces exchanged between the nanopore and the monomers without success due to a bug in the ESPResSo simulation package (which I now know how to avoid). In many cases it would have provided valuable information. In Chapter 6 the friction should decrease with increasing pre-stretching force for example. The way in which the temperature and the monomer radial distribution *away* from the pore axis affects the total friction is an open question I am quite interested in.

I was fortunate in Chapter 3 to obtain the correct scaling from the simulations. This may be because the dominant scaling was N^2 and the tube geometry considerably reduced the frictional effect.

My investigation of semiflexible polymers in Chapter 8 would greatly benefit from frictional profiling. Whether friction is reduced or increased by polymer rigidity is not trivial. On the one hand, increasing rigidity reduces small length scale fluctuations on the polymer contour, which reduces the effective corrugations. On the other hand a stiff polymer which is not aligned with the pore would need a considerable torque to pass through—probably resulting in an increased friction.

Moreover, I performed fits in Chapter 8 as a means of extracting the number of monomers on the *trans*-side that need to move in order to permit the passage of monomers through the pore. However, in the theory I neglected to account for pore frictional effects. I strongly suspect that the fits to the theory are tainted by frictional contributions.

Frictional effects can be explored as an effort to slow down experimental translocation. Simulations of varying pore diameter demonstrate that the friction can be increased, yet this approach is limited to the smallest bead diameter—if the nanopore is smaller than the bead diameter translocation cannot occur. It would be quite interesting to increase the friction without involving the nanopore diameter, perhaps via a confining geometry or an attractive force between the inner pore wall (or membrane) and the polymer.

I observed in Chapter 3 that the effective friction can be reduced by introducing a long thin confining tube on the *cis*-side. If the converse is true, then the other extreme of a “flat-disc” aspect ratio briefly explored in Chapter 4 may *increase* the friction. In a similar line of thinking, perhaps the nanopore can be placed on the *side* of the channel rather than at the end.

9.1.3 Full force field

There are considerable gains in increasing the *finesse* of the LD model with little associated computational cost. The effect of the full electric field lines can be approximated via an imposed static force field across the simulation volume—not to be confused with full electro-hydrodynamics [29]. A space-dependent force field over a range of the simulation box can replace the field localized in the nanopore approximation with little computational overhead.

Hendrick de Haan’s group has started to use this approach and found the resulting initial conformations to be greatly affected [30]. Future work on modulating the initial conformations such as contained in this thesis should consider the field lines outside the nanopore using methodologies similar to reference [30]. Deformation effects on both sides of the membrane are nontrivial. These effects should be accounted for in future work on highly-driven polymer translocation.

A

APPENDIX: Coarse-grained simulations of highly driven DNA translocation from a confining nanotube

David Sean, Gary W. Slater

As the first prize winner of the 2016 CAP Best Student Oral Competition, I was invited to submit an extended abstract.

Published in *Physics in Canada*
David Sean, Gary W. Slater: Coarse-grained simulations of highly driven DNA translocation from a confining nanotube. Physics in Canada. 2016, **72**. 142–144.

The winners of the 2016 CAP Best Student Presentation Competition at the CAP Annual Congress, 2016 June 13-17, in Ottawa are listed on pages 107-108. The extended abstracts of those winners of the CAP prizes who submitted them for publication are reproduced below. Ed.

COARSE-GRAINED SIMULATIONS OF HIGHLY DRIVEN DNA TRANSLOCATION FROM A CONFINING NANOTUBE

BY DAVID SEAN AND GARY W. SLATER

Consider this brain teaser: How do you to pass a large knitted blanket to your cold friend waiting in a locked room where the only possible passage is a small keyhole in the door? The answer: simply find one of the two ends of the yarn and slowly thread it through the keyhole. As the blanket unravels on one side of the door it can be reconstructed on the other.

Entropy ensures that a long DNA chain stays as a random disordered mess in a liquid. Passing a long polymer like DNA from one side of a membrane to the other via a nm-scale hole can at first seem impossible. But much like the passage of the knitted blanket through the keyhole, a large DNA molecule can realize the “impossible” by simply unraveling itself. Since DNA is charged in solution, this process can be driven via an electric field.

Imagine now that salt ions are present in the solution. These charged ions will also be driven to transit through the small opening, resulting in an ionic countercurrent that can be measured. Since the dividing membrane is electrically insulating, the electric field lines converge at the small hole and the measured conductivity is extremely sensitive to what is happening near the so-called *nanopore*. The passage of the DNA molecule will impede the ionic passage for the duration of the DNA *translocation* event. In the lab, this can be observed as a sudden drop in the electrical conductivity between the two chambers. Experimental translocation data is typically extracted from the duration and amplitude of a recorded drop in the measured current readout [1]. The duration of this current blockage can provide information on the molecule such as its total curvilinear length. In this short report, we present computational work on using DNA translocation for applications where the focus is on determining the length of a piece of DNA.

SUMMARY

We study the effect of pre-confining DNA in a small tube prior to driven translocation. Computer simulations are presented together with a theoretical Tension-Propagation model.

There are many applications which require finding the length of the DNA molecules present in an unknown sample. For example, forensic DNA *fingerprinting* works by breaking a long DNA molecule into smaller fragments using restriction enzymes that cleave it at sequence-specific sites. The length distribution of these smaller DNA fragments constitutes the fingerprint. Traditional *macroscopic* sizing techniques relies on the separation of these smaller pieces at a *population*-level. That is, a single DNA molecule may need to be amplified multiple times in order to create a population of DNAs which will later be fragmented and size-separated.

Nanopore translocation techniques, as described above, can be integrated into portable lab-on-a-chip devices. These can be much faster, cheaper and easier to operate than traditional methods—and could potentially only require a single DNA molecule!

Oddly, with a sample solution where all the DNA molecules are identical in size, experimental distributions of translocation times are found to be surprisingly wide. The reason for this is an example of *molecular individualism*, an expression coined by Pierre-Gilles de Gennes [2].

Figure 1a depicts a handful of possible DNA shapes—or *conformations*—at the moment when the translocation process begins (the first monomer is inside the pore). As is obvious from the figure, we are using here a rather coarse-grained representation of DNA—a simple chain of beads and springs. The generic polymer model of $N = 100$ beads is described in more details in ref [3]. The focusing of the electric field inside the pore completely dwarves its effect outside the pore. This permits us to simplify the problem further: the effect of

David Sean,
<david.sean@
gmail.com>,

and

Gary W. Slater,
<gary.slater@
uottawa.ca>,
University of Ottawa,
150 Louis-Pasteur,
Ottawa,
ON K1N 6N5

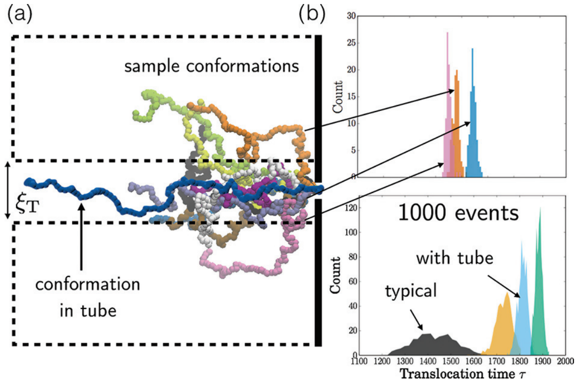


Fig. 1 (a) Placing a tube (depicted as dashed lines) removes many of the conformations found in the absence of such a tube. (b) Simulation histograms of the translocation time. Top: Three tube-free cases obtained by using the same initial conformations for multiple events. Bottom: results arising from different conformations using a typical translocation setup (no tube) and results from using three different infinitely long tubes.

the driving field is modeled as a force only applied to the beads present inside the nanopore.

We show in Fig. 1b the corresponding histograms of translocation times. The top panel is generated by repeating the simulation with the same conformation, but with different Brownian trajectories. A remarkable feature is how the different distributions are centred about their own mean translocation times. Since the translocation process occurs much faster than the time it takes for the polymer to relax (typically by two orders of magnitude), the exact positions of all the beads at the onset of translocation is a determining factor in the resulting transit time. In the lab, one cannot explicitly control for the shape of the DNA molecule once the first bead enters the pore. The resulting histograms contain a large mix of possible shapes yielding a rather wide distribution of translocation times for what is in essence a group of chemically-identical molecules! Indeed, the bottom panel of Fig. 1b shows how wide the distributions can become. Thus one way to tighten them would be to somehow limit the range of initial DNA conformations. In Fig. 1a there is a schematic of a confining tube which we use as a way to limit the possible initial conformations. We will assume that this hypothetical tube is made of a porous material such that it does not significantly affect the flow of ions or distort the field lines.

We report here the use of long (longer than the polymer) confining tube characterized only by its diameter ξ_T. By construction, the tube will limit the range of possible conformations, which by virtue of our preceding argument, should reduce the width of the distribution of translocation times.

We first generate a DNA conformation that has its first bead inside the nanopore. The presence of the tube imposes a hard cutoff $\rho = \xi_T/2$ in the radial positions of the remaining beads. For entropic reasons, a polymer chain does not like to be compressed. When we impose radial restrictions, it reacts by swelling outwards in the axial direction. As a side-effect, a thin tube will tend to squeeze out the DNA away in the axial direction. This means that the remaining monomers will on average be positioned further from the pore and will need to travel longer to translocate to the other side, as shown in the insets of Fig. 2a.

Although it may seem that this "distancing" effect alone should yield longer translocation times, there is another—perhaps subtler—contributing factor relating to the translocation *rate*. Let us investigate this according to what is known as the Tension-Propagation Theory [4].

In these out-of-equilibrium dynamics of translocation, the driving force causes a *tension* that propagates down the polymer (see red beads in Fig. 2b). If we neglect the crowding of monomers on the *trans*-side and friction in the nanopore—which contribute to minor corrections—a force-balance argument can be made to show that the instantaneous translocation rate is proportional to the number of monomers that feel this tension.

Consider the forces acting on the bead inside the pore in Fig. 2b. In the overdamped limit, the driving force F_d applied to this bead is balanced by the drag force of all moving beads under tension (coloured in red). Every bead contributes to a viscous force $-\zeta v$, with ζ the bead's friction coefficient and v the velocity. If there are k beads moving together as shown in Fig. 2b (red beads), then the total friction resisting the applied force is $-k\zeta v$ which enables us to find the translocation rate (or terminal velocity) of $v = F_d/k\zeta$. Since the driving force F_d and the friction coefficient ζ are known quantities, the resulting rate can be determined for every step of the process if we can find how many beads are affected by the applied force, i.e., how tension propagates.

Note that the beads outside the tension front do not (yet) contribute to this resistance. In fact, these beads do not even know that translocation has started. Since we consider the limit where the process is highly driven, a geometrical argument can be made to estimate the number of monomers under tension once the tension front reaches a specific bead. Figure 2b shows how the initial position of the i^{th} bead can be used to estimate the number of beads under tension once the tension front reaches it. Since the monomers under tension form a taught segment, the distance R_i between the i^{th} bead and the pore can be used to find the number of moving monomers (and the translocation rate) once the tension front reaches it.

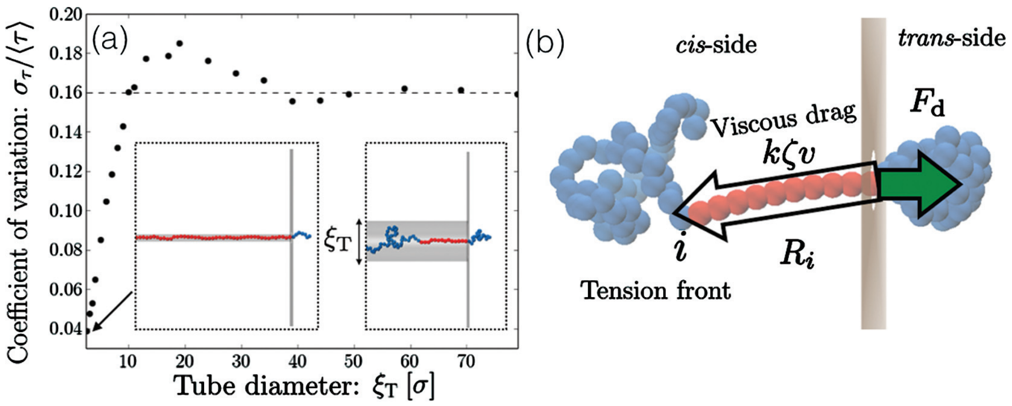


Fig. 2 (a) The coefficient of variation $\sigma_\tau/\langle\tau\rangle$ plotted for different tube radii. (b) A free-body diagram centred on the bead inside the nanopore. The drag from the beads under tension (red) opposes the force applied in the pore F_d. The i^{th} bead's initial distance R_i is used to find the amount of beads under driven motion (red) once the tension front has reached it.

This picture highlights why the initial conformations matter so much in determining the total translocation time: the initial positions essentially determine what values of viscous damping is felt by the polymer during translocation. Thin tubes will yield longer translocation events because it also causes the system to sample what is effectively a higher amount of viscous damping. Imagine pulling a heavy 1 km long chain lined-up straight on a road, versus pulling a similar chain but which is instead neatly coiled in a pile. Displacing the lined-up chain would require moving *all* the links in unison, whereas in the coiled version, you would be able to move the chain tip for a fair amount before the accumulated drag would become overwhelming.

The total translocation time τ can be determined by integrating the rate. This is a two-step process: there is the time needed to propagate to the last bead and a time for the final retraction[3]. Combining them gives a total translocation time which can be written as

$$\tau = \frac{\zeta}{F_d}\int_0^N R_i \, di,$$

where the information in the initial conformation is completely captured by the list of initial monomer distances R_i (see Fig. 2b). Averaging the above, we obtain the mean translocation time $\langle\tau\rangle = \zeta N\langle R\rangle F_d$. Note that small tube diameters not only reduce the standard-deviation σ_τ because they limit conformations, but also *increase* the mean translocation time $\langle\tau\rangle$ by pushing the beads away from the pore and increasing the friction. These two effects work together in reducing the coefficient of variation $\sigma_\tau/\langle\tau\rangle$ four-fold, as shown in Fig. 2a. When using a given mixture of multiple DNA fragments, decreasing the coefficient of variation $\sigma_\tau/\langle\tau\rangle$ means that DNA sizing can be obtained a higher resolution.

REFERENCES

1. C. Dekker, “Solid-state nanopores”, *Nat. Nanotechnol.*, **2**, 209-215 (2007).
2. P.G. deGennes, “Molecular Individualism”, *Science*, **276**, 1999-2000 (1997).
3. D. Sean, H.W. de Haan, and G.W. Slater, “Translocation of a polymer through a nanopore starting from a confining nanotube”, *Electrophoresis*, **36**, 682-691 (2015).
4. T. Saito and T. Sakaue, “Process time distribution of driven polymer transport”, *Phys. Rev. E*, **85**, 061803 (2012).

B

APPENDIX: Can gel concentration gradients improve two-dimensional DNA displays?

David Sean, Yixuan E. Wang, Gary W. Slater

Published in *Electrophoresis*
David Sean, Yixuan E. Wang, Gary W. Slater: Can gel concentration gradients improve two-dimensional DNA displays?. Electrophoresis. 2014, **35**. 736-745.

http://doi.org/10.1002/elps.201300412

Electrophoresis 2014, *35*, 736–745

David Sean
Yixuan E. Wang
Gary W. Slater

Department of Physics,
University of Ottawa, Ottawa,
Ontario, Canada

Received August 21, 2013
Revised November 11, 2013
Accepted December 9, 2013

Research Article

Can gel concentration gradients improve two-dimensional DNA displays?

The abrupt reduction in gel electrophoretic mobility that is observed when a dsDNA fragment is partially denatured has recently been predicted to exhibit a dependence upon the gel pore size. Using theoretical modeling, we demonstrate that this dependence can be exploited and used to improve the performance of 2D display of DNA. We report experimental evidence of this dependence and propose a new separation system in which a gel porosity gradient is utilized in a way analogous to temperature or denaturant gradients in traditional 2D display. Such gel porosity gradients can also be used in conjunction with denaturant gradients to improve 2D display results. We test these new ideas by modeling the fragment mobilities and computing the final fragment positions to find optimal 2D separation conditions.

Keywords:
Comparative genomics / Denaturing gradient gel electrophoresis / DNA gel electrophoresis / Electrophoretic mobility modeling DOI 10.1002/elps.201300412

1 Introduction

The hydrogen bonds which bind two complementary bases—and ultimately two ssDNA strands—can be destabilized by introducing a chemical denaturant (such as formamide) [1–3] or by increasing the surrounding temperature [4–7]. The significant cooperativity between adjacent basepairs can lead to a catastrophic event where multiple basepairs are broken as a group forming a continuous denatured region (a bubble or a split-end) [8]. Partial denaturation significantly changes DNA topology and, in turn, can profoundly impact the gel electrophoretic mobility of the DNA molecules: in fact, partially melted DNA molecules can become effectively blocked, or trapped, during gel electrophoresis [9].

Fischer and Lerman [10] were the first to use denaturant gradient gel electrophoresis (DGGE) to analyze DNA. In the first dimension of this 2D method, DNA digests are separated by size via standard gel electrophoresis. The direction of the field is then rotated by 90° and the fragments are electrophoresed in a gel containing an ascending gradient of either temperature (TGGE) or denaturant (DGGE). As the fragments migrate toward regions favoring denaturation, they eventually reach a point which provokes a topological change causing them to stop migrating. Since the probability of denaturation is composition-specific (GC basepairs are more stable than AT basepairs), sequence information (such as the presence of one or several mutations) can be inferred from the final position of the DNA molecules. One of the issues with this method though is the fact that only a fraction of the gel surface area is used [11]. In this manuscript, we propose to use pore size gradients as a new approach to increasing the gel coverage.

Lerman et al. [9] argued that there is a fundamental resemblance between the topology of a partially denatured dsDNA fragment and a branched polymer. Inspired by de Gennes' theory for the entropy-limited reptation of star polymers [12], Lerman et al. suggested that the gel electrophoretic mobility of partially denatured dsDNA should exhibit an exponential decrease with the total number of denatured basepairs p:

$$\mu = \mu_0 e^{-p/L_\mathrm{r}}, \tag{1}$$

where μ_0 is the electrophoretic mobility of the native dsDNA molecule in the gel, and L_r can be interpreted as the minimum number of basepairs that must be denatured in order to observe the blocking behavior. The entropic origin of the blocking behavior described by the semiempirical Eq. (1) is discussed in detail in [13]. In TGGE (or DGGE), a gradient of temperature (or denaturant) is used to make p a function of position in (the second dimension of) the gel which is then used to achieve sequence-based separation in the orthogonal direction.

Essentially all experimental papers making use of Eq. (1) have so far treated L_r as a constant, and most have actually assumed that its value is solely related to the properties of DNA. In de Gennes' lattice gel model, however, the branch units are measured in terms of the number of filled pores (and not in number of monomers or bases), suggesting that

Correspondence: Professor Gary W. Slater, Department of Physics, University of Ottawa, 150 Louis- Pasteur, Ottawa, Ontario, K1N 6N5, Canada
E-mail: gary.slater@uOttawa.ca
Fax: +1-613-562-5190

Abbreviation: DGGE, denaturant gradient gel electrophoresis

Colour Online: See the article online to view Figs. 1–6 in colour.

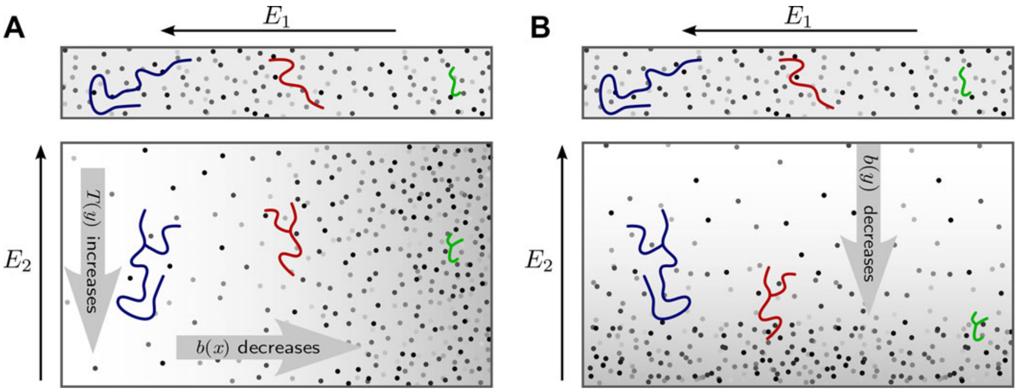


Figure 1. Schematic of the two step separation process in the presence of a pore size gradient. Once the fragments are size-separated in the first dimension (with a uniform gel and a uniform temperature), a second electric field is applied that forces them to migrate down a different gel. (A) The second gel has a pore-size gradient in the *x*-direction and a temperature gradient in the *y*-direction. (B) The second gel has a pore-size gradient in the *y*-direction and a uniform temperature.

L_r should depend on the mean pore size. In fact, a branch filling only a few pores was predicted to be enough to impact the reptation of an otherwise linear polymer chain. Therefore, as we recently demonstrated, L_r should be interpreted as the number of denatured bases required to fill a gel pore. More precisely, our theory predicted that L_r should scale with pore size b as [13]:

$$L_r \sim b^2, \tag{2}$$

which is the scaling law expected for an ideal random coil polymer chain whose persistence length is small compared to both L_r and b. We thus propose the following new concept: instead of making p position-dependent, one could use a pore size gradient to make L_r position-dependent. Indeed, we argue that this should lead to equivalent 2D separation results since it is the argument of the exponential in Eq. (1), p/L_r, that must vary along the direction of migration to achieve separation. Somewhat similar problems have been investigated where DNA is attached to a sphere [14–17], but as far as we know, this simple concept has never before been explored or proposed for denatured DNA.

In this manuscript, we provide a proof-of-concept demonstration of ways in which the dependence of L_r upon the gel pore size b can be exploited, (see Fig. 1). We give examples in which a gradient of porosity can aid to increase the spreading of the DNA fragments in 2D displays using TGGE/DGGE. We also show that a gradient in gel pore size can actually be used as the main control parameter which can ultimately replace temperature or chemical gradients and lead to new technologies. It is well established that gels can be cast containing pore size gradients [18–20]. Precast gels with a linear [21] or exponential [22] gradients can be readily purchased. This opens the door to finer control over separation conditions as temperature (or denaturant) gradients are linear by nature.

We also present a metric to evaluate the performance of 2D display methods. Being able to use different types of gradients, with different shapes, simultaneously in two directions, increases the number of ways a separation method can be customized for specific purposes. This article aims to stimulate the development of novel methods based on the general ideas presented therein.

2 Materials and methods

In this proof-of-principal demonstration of the potential usefulness of porosity gradients in 2D display of DNA, we model the separation of λ-DNA digested with *Eco*RI, *Eco*47I, *Eco*91I, *Hind*III, and *Pst*I restriction enzymes. These "textbook" fragments span the range of 15–23 130 bp. By choosing these fragments as working examples, we provide concrete examples for the novel ideas we propose here, and we can qualitatively compare our optimized DNA displays to results published in the literature. In practice, a gel size must be chosen; we chose to do our calculations for a gel of size $x_{\max} \times y_{\max} =$ 50 cm × 50 cm. Fragments that migrate past these limits are considered lost. To numerically predict the electrophoretic outcomes, we approach the problem in a way somewhat similar to what was used in [23, 24]. We repeat here the general ideas as well as specific points which need to be addressed when considering porosity gradients.

2.1 First dimension: Separation by size

Once the size and sequence of the DNA fragments are determined, the gel electrophoretic mobilities in the first dimension (x) are computed using the Wan Winkle formulae which interpolates between the limiting mobilities for electrophoresis conducted in agarose gels. The gel electrophoretic mobility μ_0 of a native (i.e. not denatured) fragment of size M (in basepairs) is found using the empirical expression [25,26]:

$$\frac{1}{\mu_0(M,c)} = \frac{1}{\mu_l} - \left(\frac{1}{\mu_l} - \frac{1}{\mu_s}\right) e^{-M/m} \tag{3}$$

where c is the agarose gel concentration (in percent). The asymptotic limiting mobilities of a "long" $\mu_l(c)$ and a "short" $\mu_s(c)$ molecule as well as the characteristic molecular size

$m(c)$ which separates these two limiting cases are given by [23, 26]:

$$\mu_l(c) = \left(1.99e^{-1.59\times c}\right) \times 10^{-4}\ \text{cm}^2/\text{Vs} \quad (4)$$

$$\mu_s(c) = (3.56 - 0.58c) \times 10^{-4}\ \text{cm}^2/\text{Vs}, \quad (5)$$

$$m(c) = (7.49 + 2.78c) \times 10^3\ \text{bp}. \quad (6)$$

Given the number of basepairs M in a fragment, Eqs. (3)–(6) will yield the fragment's native mobility $\mu_0(M, c)$ for various agarose gel concentrations c. The final x-positions of the DNA fragments (the top gels in Fig. 1) are then given by

$$x(M) = E_1 t_1 \mu_0(M). \quad (7)$$

where t_1 and E_1 are the time duration and field intensity used for the initial separation along the x-direction.

In all the examples shown in this manuscript, the results of the first separation step are calculated under identical conditions: we use $c = 0.75\%$, $E_1 = 4$ V/cm, and $t_1 = 9$ h, which yields a distribution of fragments in the gel in the approximate range $x = 20$ to 50 cm. Once the native fragments are separated by size, the field is rotated by 90° to drive the fragments into the second gel where partial denaturation can take place.

2.2 Fragment denaturation

Partial denaturation of dsDNA fragments can be achieved by raising the temperature. This is the approach (TGGE) that we will use for the examples below (please note, however, that we will neglect the effect of this increased temperature on the viscosity of the buffer). DGGE uses a chemical denaturant to achieve a similar result; it is possible to make the connection between the two approaches by converting a given concentration of chemical denaturant to an effective temperature [27]. To alleviate the text, we will refrain from using the dyad TGGE/DGGE and simply refer to TGGE.

2.3 Second dimension: Separation by topology

In the second dimension (y), topological changes in the fragment are initiated by partial denaturation (increased temperature). Since the sequence of the digests are known, a denaturation profile for each fragment is computed ahead of time. We used Meltsim [8], an open-source implementation of the Poland–Sheraga algorithm [28, 29]. Using a sodium ion concentration of 0.075 M, the denaturation profile for every fragment is precomputed and saved as a lookup table which returns the number of melted basepairs $p(T)$ at a given temperature which may be position-dependent (i.e. $T = T(y)$). For simplicity, we neglected complete strand dissociation; although it is possible to include chemical clamps (Mechlorethamine) as crosslinkers [30], these would also change the predicted melting profiles.

With the amount of denatured basepairs known for any given temperature, the Lerman equation (Eq. (1)) predicts the effect of denaturation on the mobility. Therefore, the dependence of the mobility upon the control parameters M and T can be explicitly written as:

$$\mu(T, M) = \mu_0(M)e^{-p(T)/L_r}. \quad (8)$$

Since there is no analytical form for the number of denatured basepairs as a function of position, $p(T(y))$, a fragment's velocity must be numerically integrated in order to determine the final y-position:

$$y(M, T) = E_2 \int_0^{t_2} \mu_0(M)e^{-p(T(y(t)))/L_r} dt, \quad (9)$$

where E_2 is the electric field and t_2 is the duration of the second (orthogonal) separation process. For the second separation step, we chose a time t_2 for which all of the final velocities have reduced below an arbitrary tolerance factor of 1 cm/h.

In 2D display using TGGE, temperature gradients in the second dimension (y) are created by defining a local temperature $T = T(y)$. A linear gradient of this type is thus defined using the variables T_i and T_f, the initial and final temperatures at positions $y = 0$ and $y = 50$ cm, respectively. This implicitly translates to a (nonlinear) y-dependent amount of denaturation $p(y)$ for every fragment.

Finally, for simplicity but also because we are not aware of any empirical functions equivalent to those of Van Winkle for polyacrylamide, we will assume that the migration of the native DNA fragments can also be modeled by Eqs. (3)–(6) in the second dimension.

2.4 Second dimension: Porosity gradients

The introduction of porosity gradients is the central novelty of this article. Linear gradients of pore size b will be defined in a manner similar to temperature gradients, i.e. they are defined using the initial and final pore sizes b_i and b_f. Gradients pointing along different directions (x and y) will be investigated (see Fig. 1). When used as a substitute for temperature gradients, the pore size b is chosen to be dependent upon the y-position; that is, $b = b(y)$. Alternatively, the pore size b will change along the x-direction when used to optimize the results of the TGGE process.

When considering the second dimension separation in a scenario with a position-dependent pore size, the terms in Eq. (9) must be considered carefully. Although we want to exploit the pore-size dependence of the blocking parameter L_r, we must keep in mind the fact that the native mobility μ_0 is also a strong function of the gel concentration. To make the link between the gel concentration c and the mean gel pore size b, we use the relation $b(c) = 89c^{-0.65}$ nm proposed for agarose gels [31, 32]. We thus recast the Van Winkle formulae

Eqs. (3)–(6) to contain an explicit dependence upon $b(c)$ (given in nm):

$$\frac{1}{\mu_0(M,b)} = \frac{1}{\mu_l} - \left(\frac{1}{\mu_l} - \frac{1}{\mu_s}\right) e^{-M/m} \tag{10}$$

$$\mu_l(b) = 1.99 \exp\left(-1.59\left[\frac{b}{89}\right]^{-1.54}\right) \times 10^{-4}\ \mathrm{cm^2/Vs} \tag{11}$$

$$\mu_s(b) = \left(3.56 - 0.58\left[\frac{b}{89}\right]^{-1.54}\right) \times 10^{-4}\ \mathrm{cm^2/Vs}, \tag{12}$$

$$m(b) = \left(7.49 + 2.78\left[\frac{b}{89}\right]^{-1.54}\right) \times 10^{3}\ \mathrm{bp}. \tag{13}$$

Using the pore size dependence of $L_r = L_r(b)$ as written in Eq. (2) (the prefactor will be discussed below) with $\mu_0(M,b)$ calculated from Eqs. (3)–(6), the mobility's dependence upon the mean gel pore size b can explicitly be written as:

$$\mu(M,b) = \mu_0(M,b) \times e^{-p/L_r(b)}, \tag{14}$$

where $b = b(x,y)$ is a some function of the fragment's position in the gel. We will test a range of values of b between 20 and 150 nm.

2.5 Gel coverage factor

To quantify the performance of the different systems proposed in this paper, we need a metric which reflects what is desired in a good 2D separation process. Generally speaking, for the display to be optimal the points (molecules) must maximize the use of the gel surface area. As mentioned previously, we keep the gel size constant at $x_{\max} = y_{\max} = 50$ cm. It can sometimes happen that fragments migrate outside these limits. A good metric needs to also include penalties should fragments exit the gel.

We found χ^2, the sum of the square distances amongst all N points normalized by the gel area, to work well for our needs:

$$\chi^2 = \left(\frac{1}{x_{\max} y_{\max}}\right) \sum_{i<j}^{N}{}' r_{ij}^2. \tag{15}$$

In this expression, the square distance between fragment i and j, defined as r_{ij}^2, is summed over all unique pairs ($i < j$ insures that we do not double count the distance $r_{ij}^2 = r_{ji}^2$). This sum is made dimensionless by normalizing it with the total gel area ($x_{\max} y_{\max}$). The $'$ indicates that the summation is only conducted over the N active fragments—fragments that have moved over at least 1 cm but which have not migrated beyond the gel limit. Since the end result of the summation depends on the number of active fragments, there is a built-in penalty for inactive fragments. Maximizing χ^2 (which is dimensionless) was found to provide an efficient way to characterize the performance of the various schemes tested in our study.

Table 1. Some published L_r values with the corresponding polyacrylamide gel concentrations and estimated pore size (the latter were calculated using the equations found in [39])

Pore size b (nm)	L_r (bp)	%T	References	Molecules
80	75	8	Abrams and Stanton 1992 [33]	DNA + GC clamps
89	60	6.5	Zhu and Wartell 1997 [34]	RNA
89	85	6.5	Myers et al. 1985 [35]	DNA + GC clamps
89	100	6.5	Ke and Wartell 1995 [36]	DNA
114	40	4	Mercier et al. 2008 [23]	DNA
118	40	8.13	Steger 1994 [37]	DNA + GC clamps
135	150	5.12	Riesner et al. 1989 [38]	RNA
135	200	5.12	Steger 1994 [37]	RNA

3 Results and discussion

3.1 Pore size dependence of L_r

It has previously been speculated that L_r should be a function of the mean gel pore size [34, 37, 40]. Unfortunately, many factors can influence the value of L_r and our attempts to identify trends between the values of L_r published in the literature and the corresponding gel pore sizes (see Table 1) are inconclusive. Furthermore, the use of RNA molecules instead of DNA was observed to systematically affect L_r [34, 37].

Another complication may arise when using chemical or GC clamps to prevent complete strand disassociation. Situations where GC clamps are used [37,41,42] may systematically favor denaturation regions of the bubble-variety. This in itself is not a problem as Lerman et al. suggested, a bubble may form a branch unit which will hinder migration [9]. However, computer simulations demonstrate that the values of L_r are systematically higher for denaturation in the form of bubbles than at the fragment ends [43]. Thus situations which use GC clamps can favor the creation of bubbles would inevitably affect L_r.

It has been clearly demonstrated (see for instance the Rosenbaum and Klahn photographs published by Riesner et al. [40]) that the blocking behavior is more pronounced for high electric fields. In that study, RNA was electrophoresed in gels containing a perpendicular temperature gradient under different voltages (100, 300, 400, and 600 V). It was found that the higher voltage resulted in a greater contrast between the RNA's native and partially denatured mobilities. This is in qualitative agreement with the fact Langevin Dynamics simulations probing the effect of increasing the electric field for split-end DNA predict that high fields should promote hooking interactions and/or cause the molecule to collapse into metastable conformational traps [44]. Therefore, both the Riesner et al. [40] experimental results and our simulation study [44] show that the field strength is yet another factor to account for.

When these complications are considered, an assessment of the pore size dependence of L_r extracted from the literature is infeasible. As far as we know, there is no published systematic experimental study of the parameter L_r. However, in addition to testing for field dependencies, Riesner et al. also investigated the blocking behavior using three different polyacrylamide gel concentrations [40]. Between the three concentrations tested (6, 8, and 10%), a higher contrast of mobility is observed for the 10% gel; i.e. a lower L_r is observed for a gel with smaller pores. This indirect observation qualitatively agrees with the behavior predicted by Eq. (2). We extracted the ratio μ/μ_0 from the published photographs and used $\ln(-\mu/\mu_0)$ to estimate p/L_r. Since we did not know the value of p for these experiments, the prefactor k in:

$$L_r = kb^2, \tag{16}$$

was determined by fitting the data points with the unkown value of p absorbed into k. We motivated the use of a slope of 2 here by the theoretical scaling prediction of Eq. (2). We do not expect numerical consistency between these RNA L_r and ones for λ-DNA. However, based on theoretical grounds we expect the scaling with pore size to be the same. The mean pore size is estimated using $b = 231(\%T)^{-0.51}$ [39]. In Fig. 2, we plot the results against the pore size and correct for the prefactor by rescaling the y-axis by the fitted value k.

Additionally, Bénédicte Lafay (Laboratoire Ampère, École Centrale de Lyon) generously provided us with TGGE photographs of λ-DNA digested with *Eco*RI. These included six datasets in which the experimental conditions were identical, save the value of the polyacrylamide gel concentration $\%T$ and some irrelevant differences in the duration of the electrophoresis. The second separation process for $\%T = 3$, 4, 5, 6, 7, and 8% gels ran for $t_2 = 24$, 24, 30, 40, 30, and 40 h, respectively. Using the approach described in [23], we fitted the relative positions extracted from the photographs to our model using L_r and t_2 as fitting parameters for each gel (while preserving the relative time differences). Again, we estimated the pore size using $b = 231(\%T)^{-0.51}$ [39]. The extracted L_r values were fitted using Eq. (16) to obtain the empirical prefactor k (in this case, the value of p is known). When fitting this data, the outlier point corresponding to the largest pore size was excluded. For these λ-DNA digests, $k = 0.003$ bp nm^{-2} is obtained from the fit shown in Fig. 2. We use this empirical prefactor in Eq. (16) to model L_r in the remaining calculations. Overall, the b^2 scaling law appears to be consistent with the limited amount of information provided by these two datasets.

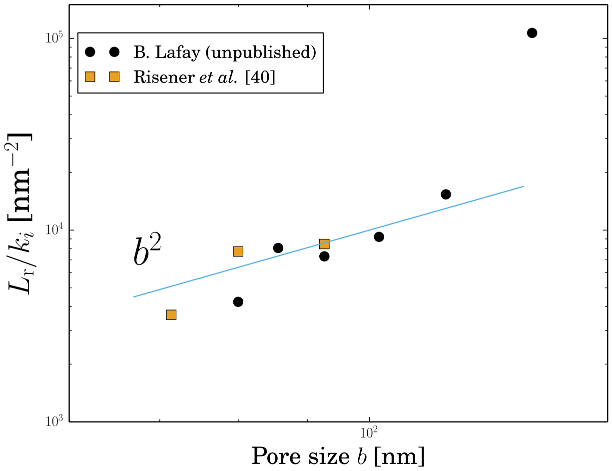


Figure 2. The ratio L_r/k_i is plotted as a function of the estimated gel pore size on a log–log scale. For each dataset a fit using Eq. (16) was calculated. The extracted values of L_r were then rescaled with their corresponding fitted k_i. The outlier data point from B. Lafay is omitted from the fit.

3.2 Optimization using χ^2

The χ^2 metric as defined by Eq. (15) is first tested for a traditional TGGE 2D display. We use a uniform gel concentration of $c = 0.75\%$ as in [23] and test various temperature gradients by varying the initial and final temperatures T_i and T_f, respectively. For each gradient tested, the gel coverage is evaluated by calculating χ^2 for the final position of the separated fragments. The resulting phase space map of χ^2 is shown in Fig. 3D. We also include in this figure three examples of separation results, including the temperature gradient in [23], Fig. 3A; an optimal case, Fig. 3B; and a poorly optimized case, Fig. 3D.

The final fragment positions corresponding to the gradient conditions of [23], i.e. $T_i = 60$°C and $T_f = 100$°C, are shown in Fig. 3A. This case results in a value of $\chi^2 = 2.36$.

Using the heat map as a guide, the temperature gradient was tuned for the optimal display shown in Fig. 3B. Here, a gradient with $T_i = 74$°C and $T_f = 87$°C gives the optimal value of $\chi^2 = 7.73$.

Lastly, we present a typical "bad-case" scenario in Fig. 3C. The $T_i = 50$°C and $T_f = 84$°C gradient is found by choosing a point from the heat map which gives a low coverage of $\chi^2 = 0.88$. Visual inspection of these three cases validates the proposed factor χ^2 as a metric to characterize the performance of the separation procedure. We now proceed to examine different porosity gradients using the numerical values of χ^2 as a guide.

3.3 Porosity gradients as an alternative to temperature gradients

We now show that we can exploit the dependence of L_r upon the gel concentration and use it as a primary control parameter in the separation-by-sequence step of 2D DNA displays. Instead of a thermal gradient, which is the main actor in controlling $p(T)$ in the argument of the the exponential in Eq. (8), we choose to have a uniform temperature and rather control $L_r(b)$ via the gel pore size, viz., we choose a position-dependent pore size $b(y)$ to replace a position-dependent temperature $T(y)$. In this situation, as with most 2D display protocols, the fragments are first size-separated in the x-direction

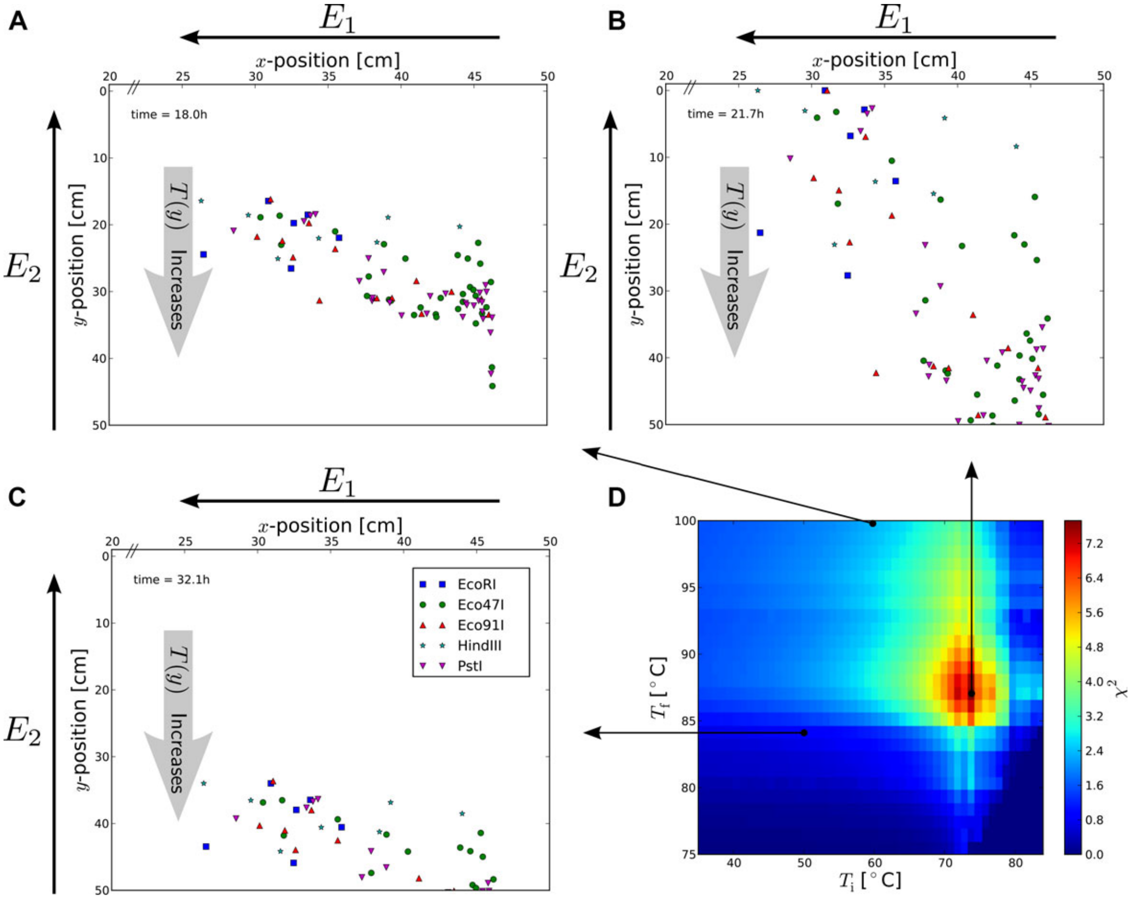


Figure 3. (Color online) TGGE examples showcasing parameter χ^2 as a metric for characterizing gel coverage. (A) The temperature gradient used in [23] gives $\chi^2 = 2.36$. (B) The optimal temperature gradient gives the maximum value of $\chi^2 = 7.73$. (C) Parameters chosen from a low $\chi^2 = 0.88$ point yields a low coverage. (D) The phase space map of χ^2 used to explore the 2D phase space of temperature gradients.

according to Eq. (7). However, during the second separation process, the denaturation conditions are uniform—but need to be strong enough to ensure nonzero values of p for the fragments. Once partially denatured, the fragments are electrophoresed through a gel with a negative porosity gradient pointing in the direction of migration (y). Using a gradient defined by parameters b_i and b_f (where $b_f < b_i$), the local pore size $b(y)$ is obtained from:

$$b(y) = b_i - (b_i - b_f)\left(\frac{y}{y_{\max}}\right). \tag{17}$$

While traveling down a lane with a descending value of pore sizes, a fragment will eventually reach a position where $p \approx L_r(b(y))$ and will block. Although this is ultimately a three parameter problem (T, b_i, and b_f), we choose a single uniform temperature $T = 75^\circ$C and vary only the two parameters defining the porosity gradient, b_i and b_f.

Figure 4 shows a phase space map for χ^2 together with the final DNA positions of the optimized gel. The optimal gradient ($b_i = 120$ nm and $b_f = 40$ nm) gives a coverage of $\chi^2 = 6.62$ which is significantly higher than the TGGE parameters of [23] ($\chi^2 = 2.36$) and comparable to the optimized TGGE separation presented in the previous section ($\chi^2 = 7.73$). This promising result suggests that porosity gradients can perform almost just as good as temperature gradients. Although the source of the blocking is the same, the final y-position of the digests are not controlled by the same mechanisms between porosity and temperature gradients. Interestingly, one can find examples for which overlapping points in the final standard TGGE display can be separated if a porosity gradient is used as an alternative to a temperature gradient.

However, it is important to realize that both factors in the rhs of Eq. (14) have a dependence upon the pore size b. The sequence-dependent separation step along the y-direction relies on the assumption that the last term eventually dominates the dynamics. Since it is an exponential function, it is reasonable to assume that this condition is met most of the time. Unfortunately, as we will argue here, this may not always be the case. Indeed, the blocking term $\exp\left(-\frac{p}{kb^2}\right)$ can sometimes be inconsequential. Since the ratio p/kb^2 has to be larger than unity for fragment denaturation to lead to blocking, one needs large values of p if the experimental conditions are such that $kb^2 \gg 1$. But large values of p require long DNA molecules

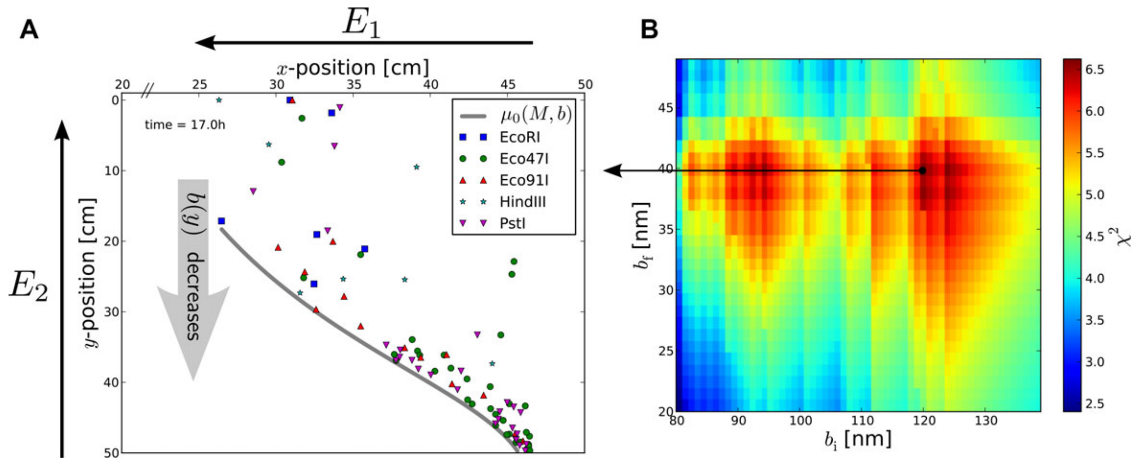


Figure 4. (Color online) 2D display using a porosity gradient along the y-direction. The empirical prefactor of L_r in Eq. (16) was used to generate the plots. (A) The final fragment positions using an optimized gel with gradient $b_i = 120$ nm and $b_f = 40$ nm gives a coverage of $\chi^2 = 6.62$. The gray line shows the predicted final positions of the fragments if there were no blocking. (B) The heat-map of χ^2 used to find the optimal gradient conditions.

since one must have $M > p$. Therefore, the value of $L_r = kb^2$ (and hence the choice of a gel) determines the range of DNA sizes that can be analyzed using pore size gradients.

Moreover, there is also an exponential dependence of the mobility of the native fragment upon the mean pore size in the long DNA limit:

$$\mu_0 \simeq \mu_l \sim \exp\left(-1.59\left[\frac{89}{b}\right]^{1.54}\right). \tag{18}$$

Shrinking the pore size b down until the blocking term dominates may simply result in a mobility μ_0 too low for any practical use. We show in Fig. 4A the μ_0 backbone which limits the maximum migration of the fragments. The feasibility of using a porosity gradient in the manner proposed here thus hinges on the numerical contrast between k and the parameters of Eq. (18). In some cases, the native mobility of the fragment may become negligible before the denatured parts of the fragment lead to trapping in the gel. Careful selection of the experimental conditions would thus be required. The ideas presented here can serve as a guide for doing so.

3.4 Perpendicular porosity gradient

2D DNA display methods share a common shortcoming in that the dispersed points are often concentrated along the upper diagonal of the gel [11]. This is a consequence of the size-dependence of the mobility of the DNA fragments: the small fragments that travel quickly during the first electrophoretic step will most likely travel quickly in the second dimension (separation by denaturation). Similarly, longer fragments will likely travel slowly during both steps. This results in a dominant diagonal about which the majority of the points are distributed (this is the gray line in Fig. 4). For the very same reason, the second electrophoretic step is more time consuming since one must wait to allow longer fragments to migrate toward zones in the gel that favor denaturation; zones that smaller fragments would have no problems reaching.

In response to the bottleneck being the migration of the longest fragments, a gel can be engineered to contain a porosity gradient where the longer fragments would travel down a lane containing larger pores. In principle, one could indeed design the gel in such a way that all fragments will migrate toward higher temperatures at roughly the same rate. Such an experimental set up would be similar to what is used for TGGE 2D displays. After size-separating the fragments (in the x-direction), the electric field is rotated by 90° and fragments are electrophoresed (in the y-direction) in a gel containing an ascending temperature gradient. The gel is cast to contain a porosity gradient in the x-direction, such that the larger fragments migrate down a lane characterized by larger pores than the lanes occupied by the smaller fragments.

This certainly changes the interpretation of the final fragment positions. With traditional TGGE, the final y-positions are determined when the number of denatured basepairs, $p(T)$, is comparable to some *constant* value of L_r. In this case, all fragments are thus upheld to the same criterion—they all stop migrating after a specific absolute number of basepairs have denatured ($p(T) \approx L_r$). When the second gel contains the described transverse porosity gradient, the fragments are blocked using a criterion that is implicitly related to their length; e.g. the smaller fragments may block when $p(T) \approx 40$ bp (if $L_r = 40$ bp in the part of the gel with smaller pores) and the longer fragments may block when $p(T) \approx 200$ bp (if $L_r = 200$ bp in the part of the gel with larger pores). This may actually increase the range of usable fragment sizes during a single separation run.

As this situation contains two gradients, there are four independent parameters to vary: T_i, T_f, b_i, and b_f. For simplicity, we first test this approach using a temperature gradient identical what was used in Fig. 3A and [23]. In addition

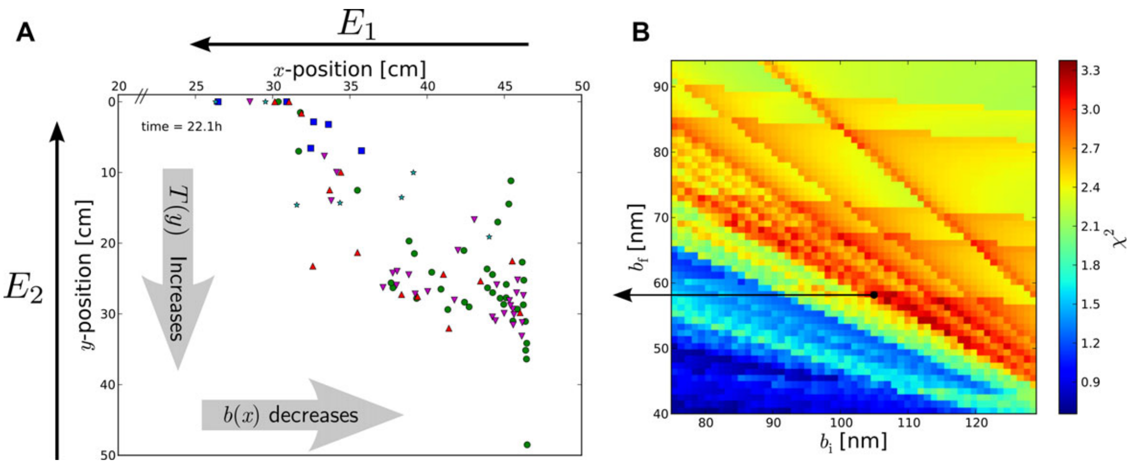


Figure 5. (Color online) In addition to using the TGGE conditions of [23], a linear porosity gradient is cast in the *x*-direction. (A) The gel coverage is optimized using $b_i = 105$ nm and $b_i = 58$ nm. (B) The χ^2 phase space map used to obtain the optimal porosity gradient.

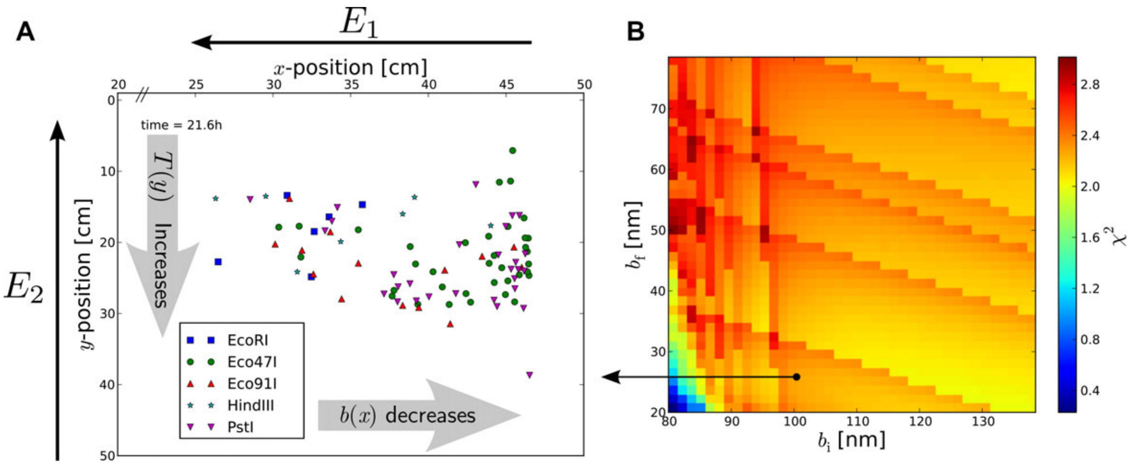


Figure 6. (Color online) In addition to using the TGGE conditions of [23], a quartic porosity gradient is cast in the *x*-direction. (A) The quartic gradient reduces the diagonal features of the final dispersion of molecules in the gel. (B) The phase space map of χ^2 for the quartic porosity gradient.

to a temperature gradient characterized by $T_i = 60^\circ$C and $T_f = 100^\circ$C, we model a porosity gradient by defining a local pore size for every fragment which depends linearly upon their x-positions:

$$b(x) = b_i - (b_i - b_f)\left(\frac{x}{x_{\max}}\right). \tag{19}$$

Figure 5A shows a map of the final positions obtained using $b_i = 105$ nm and $b_f = 58$ nm. These optimal parameters were found using the χ^2 phase space map in Fig. 5B. The resulting value of $\chi^2 = 3.38$ is higher than the previous $\chi^2 = 2.36$ obtained for a gel with a uniform pore size.

Alternatively, we can exploit the possibility of using nonlinear gradients here. We thus repeated the previous optimization using a pore size which has a much sharper quartic x-dependence,

$$b(x) = b_i - (b_i - b_f)\left(\frac{x}{x_{\max}}\right)^4. \tag{20}$$

We motivate the use of a steeper gradient by observing that the dominating diagonal—although not as severe as what we observe for the uniform pore size case shown in Figs. 3A and 4A—is still perceptible in Fig. 5A. The diagonal feature appears to be mostly eliminated by tuning the quartic gradient to $b_i = 100$ nm and $b_f = 25$ nm as shown in Fig. 6A—the smaller fragments block earlier while the longer fragments migrate further down the gel. It is worth noting here that although the diagonal effect has been reduced, the calculated value of $\chi^2 = 2.28$ is slightly lower than the linear gradient case, $\chi^2 = 3.38$. This is due to the fact that the lower portion of the gel remains unoccupied. One could in principle maximize the performance of the system by using a multiple-parameter optimization process involving all experimental conditions at once; this is beyond the scope of the present paper.

4 Concluding remarks

We recently predicted that the L_r parameter, which is traditionally used together with Eq. (1) to describe the blocking of partially denatured DNA fragments in gel, should vary with the mean gel pore size since tighter gels should block molecules more easily than open ones. In this article, we used 2D display photographs provided by B. Lafay and photographs from Rosenbaum and Klahn published by Riesner et al. [40] to provide an experimental verification of this prediction. The data appears to be consistent with the predicted scaling law $L_r = kb^2$, although more data will be required to confirm this.

Based on this observation, we proposed to use gradients of gel concentration to improve the performance of TGGE separation methods. In order to investigate various schemes, we generalized a numerical 2D display model [23, 24] to account for variations in gel porosity and we introduced a metric to quantify the performance of 2D displays by evaluating the gel coverage.

Using these novel ideas and numerical tools, we have shown that a linear gradient in pore size $b(y)$ can be used as an alternative to TGGE as the second step in 2D DNA display. While these separations perform in a way quite comparable to TGGE, blocking is controlled via an inherently different process. This new approach could be useful in situations where TGGE fails to provide adequate resolution between two fragments. Moreover, temperature or chemical gradients are typically linear by nature. Gels on the other hand can be engineered to contain nonlinear gradients or in extreme cases embedded with gels of different concentrations [45]. Manipulating the gel pore size should thus provide more flexibility to the user. However, we have also demonstrated that the experimental conditions must be chosen with great care in order to be able to obtain separations controlled by DNA denaturation.

Lastly, we proposed the use of porosity gradients as an aid to TGGE by orienting the gradient in the x-direction. Since the pore size $b(x)$ then depends on the x-position of a fragment, long fragments migrate down lanes with larger pores than small fragments. This addresses a common bottleneck in TGGE, namely, while small fragments readily migrate down to denaturing conditions, big fragments approach these positions much more slowly. We successfully tested this idea using both linear and nonlinear pore size gradients. This is clearly a simple and effective way to maximize the use of the gel surface area.

The authors would like to acknowledge the use of experimental TGGE photographs provided by Dr. Bénédicte Lafay. Y.E.W. was supported by the NSERC Undergraduate Student Research Award. D.S. is supported by the NSERC-CREATE training program in Quantitative Biomedicine. This work is supported in part by an NSERC grant to G.W.S.

The authors have declared no conflict of interest.

5 References

[1] Oros-Sichler, M., Gomes, N. C., Neuber, G., Smalla, K., *J. Microbiol. Meth.* 2006, *65*, 63–75.

[2] Zhang, Y., Wang, G., Wu, Y., Zhao, H., Zhang, Y., Sun, Z., *Adv. Biosci. Biotechnol.* 2013, *4*, 437–441.

[3] Cleary, D. F., Smalla, K., Mendonça-Hagler, L. C., Gomes, N. C., *PLoS One* 2012, *7*, e29380.

[4] Hernán-Gómez, S., Espinosa, J., Ubeda, J., *FEMS Microbiol. Lett.* 2000, *193*, 45–50.

[5] Reboleiro-Rivas, P., Juarez-Jiménez, B., Martinez-Toledo, M., Rodelas, B., Andrade, L., Gonzalez-Lopez, J., Fenice, M., *Int. J. Environ. Res.* 2013, *7*, 685–696.

[6] Gómez-Villalba, B., Calvo, C., Vilchez, R., González-López, J., Rodelas, B., *Appl. Microbiol. Biot.* 2006, *72*, 393–400.

[7] Calderón, K., González-Martínez, A., Montero-Puente, C., Reboleiro-Rivas, P., Poyatos, J. M., Juárez-Jiménez, B., Martínez-Toledo, M. V., Rodelas, B., *Bioresour. Technol.* 2012, *103*, 87–94.

[8] Blake, R. D., Bizzaro, J. W., Blake, J. D., Day, G., Delcourt, S., Knowles, J., Marx, K. A., SantaLucia, J., *Bioinformatics* 1999, *15*, 370–375.

[9] Lerman, L., Fischer, S., Hurley, I., Silverstein, K., Lumelsky, N., *Annu. Rev. Biophys. Bio.* 1984, *13*, 399–423.

[10] Fischer, S., Lerman, L., *Cell* 1979, *16*, 191–200.

[11] Malloff, C. A., Fernandez, R. C., Dullaghan, E. M., Stokes, R. W., Lam, W. L., *Gene* 2002, *293*, 205–211.

[12] de Gennes, P., *J. Phys.* 1975, *36*, 1199–1203.

[13] Sean, D., Slater, G. W., *Electrophoresis* 2010, *31*, 3446–3449.

[14] Serwer, P., Hayes, S. J., Watson, R. H., *J. Mol. Biol.* 1992, *223*, 999–1011.

[15] Serwer, P., Hayes, S. J., Moreno, E. T., Park, C. Y., *Biochemistry* 1992, *31*, 8397–8405.

[16] Desruisseaux, C., Slater, G. W., Drouin, G., *Macromolecules* 1998, *31*, 6499–6505.

[17] Ulanovsky, L., Drouin, G., Gilbert, W., *Nature* 1990, *343*, 190–192.

[18] Rainwater, D. L., Andres, D. W., Ford, A. L., Lowe, F., Blanche, P. J., Krauss, R. M., *J. Lipid Res.* 1992, *33*, 1876–1881.

[19] Vilim, V., Krajickova, J., *Anal. Biochem.* 1991, *197*, 34–39.

[20] Dunn, M. J., Corbett, J. M., *Method Enzymol.* 1996, *271*, 177–203.

[21] Moiseyev, G., Takahashi, Y., Chen, Y., Gentleman, S., Redmond, T. M., Crouch, R. K., Ma, J.-x., *J. Biol. Chem.* 2006, *281*, 2835–2840.

[22] Cho, D., Collins, M. T., *Clin. Vaccine Immunol.* 2006, *13*, 1155–1161.

[23] Mercier, J., Kingsburry, C., Slater, G., Lafay, B., *Electrophoresis* 2008, *29*, 1264–1272.

[24] Florescu, A.-M., Joyeux, M., Lafay, B., *Electrophoresis* 2009, *30*, 3649–3656.

[25] Rill, R. L., Beheshti, A., Van Winkle, D. H., *Electrophoresis* 2002, *23*, 2710–2719.

[26] Van Winkle, D. H., Beheshti, A., Rill, R. L., *Electrophoresis* 2002, *23*, 15–19.

[27] Lerman, L. S., Silverstein, K., *Method Enzymol.* 1986, *155*, 482–501.

[28] Fixman, M., Freire, J., *Biopolymers* 1977, *16*, 2693–2704.

[29] Poland, D., *Biopolymers* 1974, *13*, 1859–1871.

[30] Romero, R. M., Rojsittisak, P., Haworth, I. S., *Electrophoresis* 2013, *34*, 917–924.

[31] Holmes, D. L., Stellwagen, N. C., *Electrophoresis* 1990, *11*, 5–15.

[32] Slater, G. W., Rousseau, J., Noolandi, J., Turmel, C., Lalande, M., *Biopolymers* 1988, *27*, 509–524.

[33] Abrams, E. S., Stanton, V. P. Jr., *Method Enzymol.* 1992, *212*, 71–104.

[34] Zhu, J., Wartell, R., *Biochemistry* 1997, *36*, 15326–15335.

[35] Myers, R. M., Fischer, S. G., Lerman, L. S., Maniatis, T., *Nucleic Acids Res.* 1985, *13*, 3131–3145.

[36] Ke, S., Wartell, R., *Biochemistry* 1995, *34*, 4593–4600.

[37] Steger, G., *Nucleic Acids Res.* 1994, *22*, 2760–2768.

[38] Riesner, D., Steger, G., Zimmat, R., Owens, R., *Electrophoresis* 1989, *10*, 377–389.

[39] Holmes, D. L., Stellwagen, N. C., *Electrophoresis* 1991, *12*, 612–619.

[40] Riesner, D., Henco, K., Steger, G., *Advances in Electrophoresis* 1991, *4*, 169–250.

[41] Shiraishi, M., Oates, A., Li, X., Chuu, Y., Sekiya, T., *Biol. Chem.* 2004, *385*, 967–973.

[42] Myers, R., Fischer, S., Maniatis, T., Lerman, L., *Nucleic Acids Res.* 1985, *13*, 3111–3129.

[43] Sean, D., Slater, G. W., *Electrophoresis* 2012, *33*, 1341–1348.

[44] Sean, D., Slater, G. W., *Electrophoresis* 2013, *34*, 745–752.

[45] Serwer, P., *Anal. Biochem.* 1985, *144*, 172–178.

C

APPENDIX: Interfacing solid-state nanopores with gel media to slow DNA translocations

Matthew Waugh, Autumn Carlsen, **David Sean**, Gary W. Slater, Kyle Briggs, Harold Kwok, Vincent Tabard-Cossa

Published in *Electrophoresis*
Matthew Waugh, Autumn Carlsen, David Sean, Gary W. Slater, Kyle Briggs, Harold Kwok, Vincent Tabard-Cossa: Interfacing solid-state nanopores with gel media to slow DNA translocations. Electrophoresis. 2015, **36**. 1759-1767.

http://doi.org/10.1002/elps.201400488

Electrophoresis 2015, *36*, 1759–1767

Matthew Waugh
Autumn Carlsen
David Sean
Gary W. Slater
Kyle Briggs
Harold Kwok
Vincent Tabard-Cossa

Department of Physics,
University of Ottawa, Ottawa,
Ontario, Canada

Received October 17, 2014
Revised February 27, 2015
Accepted April 7, 2015

Research Article

Interfacing solid-state nanopores with gel media to slow DNA translocations

We demonstrate the ability to slow DNA translocations through solid-state nanopores by interfacing the *trans* side of the membrane with gel media. In this work, we focus on two reptation regimes: when the DNA molecule is flexible on the length scale of a gel pore, and when the DNA behaves as persistent segments in tight gel pores. The first regime is investigated using agarose gels, which produce a very wide distribution of translocation times for 5 kbp dsDNA fragments, spanning over three orders of magnitude. The second regime is attained with polyacrylamide gels, which can maintain a tight spread and produce a shift in the distribution of the translocation times by an order of magnitude for 100 bp dsDNA fragments, if intermolecular crowding on the *trans* side is avoided. While previous approaches have proven successful at slowing DNA passage, they have generally been detrimental to the S/N, capture rate, or experimental simplicity. These results establish that by controlling the regime of DNA movement exiting a nanopore interfaced with a gel medium, it is possible to address the issue of rapid biomolecule translocations through nanopores—presently one of the largest hurdles facing nanopore-based analysis—without affecting the signal quality or capture efficiency.

Keywords:
Agarose / Bionanotechnology / Next-generation sequencing / Polyacrylamide / Solid-state nanopore

DOI 10.1002/elps.201400488

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

During a typical nanopore experiment, an applied voltage drives a stream of ions through a narrow pore in a thin membrane separating two fluidic reservoirs. The passage of a biomolecule through the nanopore displaces a corresponding volume of ionic solution, resulting in a measurable current change. Ideally, this method would be capable of differentiating between individual bases on a DNA molecule using the current signature collected from the unique blockage volume of each base [1, 2, 4]. However, due to the rapid translocation rate of DNA through biological and sub-10-nm solid-state nanopores (roughly 1 nucleotide/μs and $>$10 bp/μs, respectively [4, 5]), it is extremely difficult to collect a sufficient number of ions for each base to overcome background noise and signals from adjacent bases [2, 6]. Thus, a critical step toward nanopore-based nucleic acid analysis is extending the time that DNA lingers in the nanoscale confinement of a nanopore (while minimizing the likelihood of backtracking due to Brownian motion), thereby enabling detection of more ions per subnanometer section of the strand. The slowing down of DNA translocation kinetics will not only allow sufficient signal averaging to resolve current differences required to sequence DNA [7, 8] but may also benefit various diagnostic applications by allowing better identification of labels for multiplexed biomarker detection [9, 10].

In pursuit of extending nanopore dwell times, researchers have modified three basic components of the nanopore system—the environment, the molecule of interest, and the nanopore itself. In terms of the environment surrounding the nanopore, both biological and solid-state nanopore researchers have experimented with reducing solvent temperature [11, 12] and increasing viscosity [4, 13, 14]. Solid-state researchers have further explored the use of alternate salts [15, 16] and the introduction of a salt concentration gradient [17, 18] or counter pressure [19, 20], achieving close to tenfold translocation rate reductions by these

Correspondence: Professor Vincent Tabard-Cossa, Department of Physics, University of Ottawa, 150 Louis-Pasteur, Rm MCD338J, Ottawa, ON K1N 6N5, Canada
Fax: +1-613-562-5190
E-mail: tcossa@uottawa.ca

Abbreviation: CBD, controlled breakdown

Colour Online: See the article online to view Figs. 1 to 4 in colour.

means [20]. However, since altering the nanopore environment influences the molecule of interest and the ions in solution equally, the gains in dwell time have generally come at the expense of other desirable attributes, such as high S/N or capture rate.

Efforts to slow translocation have also focused on modifying attributes of the molecules of interest, employing molecular motors or tethering them to mechanical probes. For biological nanopores, the addition of bulky tags to the analyte was found to diminish translocation speed [21] while enzymes [6] permitted ratcheting (also called processive translocation [7, 22–24]) and a closer look at DNA-enzyme interactions [25]. In fact, combining enzymes with an MspA nanopore [26] enabled single-nucleotide resolution [8] of ssDNA strands. Combining DNA hairpins [27–30] or avidin-bound ssDNA [31–33] with biological nanopores has facilitated slower translocation and repeated interrogation of the same molecule by physically blocking the molecule from completing its passage through the nanopore. Similar results have been achieved through the incorporation of optical tweezers [34, 35], magnetic tweezers [36], or atomic force microscope tips [37] in solid-state nanopore experiments. While these modifications to the analyte have substantially slowed translocation rates to greater than tens of μs per base, it should be noted that a significant selling point for a nanopore-based system is high-throughput and scalability to obtain rapid results at low cost. Many of these approaches inherently add layers of complexity, while the ideal nanopore device would not require such modifications [1].

A final route to slowing translocations has involved modification of the nanopore itself. For example, modifications to solid-state nanopores have included adjusting the nanopore diameter to increase friction [38–40], modulating the nanopore surface charge with laser light [41], functionalizing the nanopore with hydrogen-bonding molecules [42] or coating it with a lipid bilayer [43], and altering the properties of a pH-responsive organic film [44]. Additional methods have been proposed, including local heating of a gold layer surrounding the nanopore to stretch the DNA [45, 46] and ratcheting of nucleotide strands through introduction of a third electrode [47, 48]. In fact, researchers have investigated a three-terminal system, or field effect nanofluidic transistor, which would alter the electric field profile in the nanopore [49–51] and modulate its surface charge [52–55]. Base-by-base ratcheting using electrostatic traps in a DNA transistor has yet to be achieved, but nanopore modifications have already reduced translocation speeds by up to an order of magnitude for ssDNA [42, 53].

Recently, the use of an alternative method for slowing translocations has been reported—a copolymer mesh electrospun on the *cis* side of a solid-state nanopore [56]. An approaching DNA fragment must first pass through the mesh then electrophoretically thread through the nanopore before exiting on the *trans* side. Although a promising approach—translocation rates were reduced by up to two orders of magnitude, reaching ~2 bp/μs—the copolymer mesh employed has not yet been fully characterized and requires extensive experimental preparation time. Our work is distinct from this in two ways: (1) we employ readily available, well-characterized materials—agarose or polyacrylamide— and (2) we deposit the film on the *trans* side of the nanopore, thus keeping the DNA capture process intact. The application of the gel medium exclusively on the *trans* side of a solid-state nanopore (Fig. 1) with the objective of slowing DNA translocation kinetics also distinguishes our work from a previous publication that described deposition of agarose on either side of a biological nanopore to improve bilayer stability [57]. Our approach represents an important step toward cost-effective, rapid, and simple nanopore-based diagnostic capabilities. We are able to slow the translocation process while maintaining a high electric field strength, which is important for preserving a sufficiently high capture rate (or capture radius), ensuring detection of a larger number of ions flowing through the nanopore for a stronger signal, as well as overcoming the issue of DNA molecules backtracking within the nanopore due to Brownian motion.

In theory, no fewer than four length scales are needed to describe the movement of a charged polymer on the *trans* side of a nanopore membrane interfaced with a gel medium: the persistence length (L_p) and radius of gyration (R_g) of the polymer, the mean gel pore size (d_{gel}), and the convection length (R_c), defined as the distance from the nanopore beyond which Brownian motion dominates over the electrostatic forces (this length scale is analogous to the capture radius on the *cis* side [18]). While the interplay between these four length scales may, in principle, predict the existence of numerous translocation regimes, only a few of them are relevant when realistic experimental values are considered, as described below.

In order to estimate R_c, we can assume the electric field extends outwards from the nanopore in a radial fashion according to Ohm's law (assuming spherical symmetry) [58],

$$\vec{E}\,(r) = \frac{I_0 \vec{r}}{2\pi |r^3| \sigma} \tag{1}$$

where I_0 is the baseline ionic current, σ is the conductivity of the solution, and r is the distance from the mouth of the nanopore. From Eq. (1), we are able to determine that the electric field extends several micrometers into the *trans* side, and will therefore push translocated DNA away from the nanopore and further into the gel, making room for other molecules.

In order to determine the most effective gel pore size for the *trans* side of the membrane, we consider the behavior of DNA undergoing gel electrophoresis when the mean diameter of the gel pores, d_{gel}, is larger than, equal to, or smaller than the radius of gyration, R_g, of DNA. Firstly, $R_g < d_{\mathrm{gel}}$ produces Ogston sieving. In this regime, the DNA molecule must travel a distance larger than R_g before encountering a gel fiber, meaning that in our system it could not interact with a gel fiber and be within the nanopore sensing region simultaneously. Therefore Ogston sieving mechanisms cannot easily be exploited to slow down DNA translocation through the

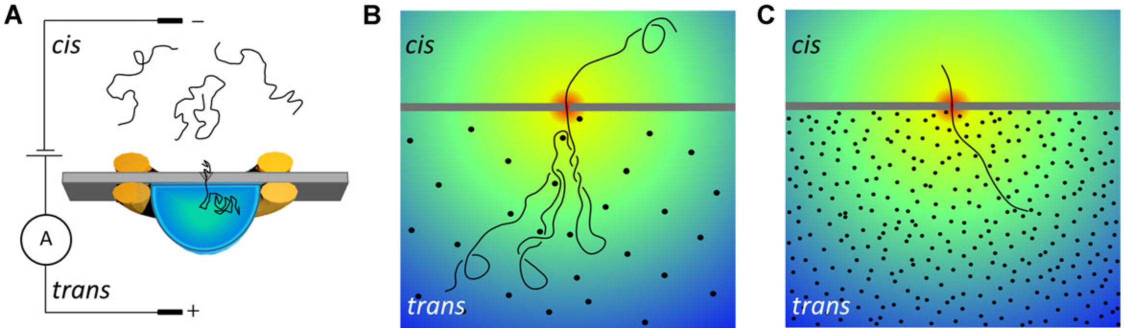


Figure 1. (A) Schematic of the experimental system (not to scale). Elastomer gaskets contain a droplet of the gel solution (agarose or polyacrylamide) deposited on the *trans* side of a standard 10-nm thick SiN_x membrane (50-μm × 50-μm) on a Si chip. Following gelation of the polymer solution and subsequent mounting of the Si chip in a fluidic cell, a nanopore is formed through the SiN_x membrane by controlled breakdown (CBD). DNA is introduced to the *cis* side of the membrane and electrophoretically driven through the nanopore into the gel medium on the *trans* side. (B,C) Schematic cross-sectional view of a DNA molecule entering into a gel medium from a nanopore, where (B) satisfies the condition $R_g > d_{\text{gel}} > L_p$, similar to the case for agarose, while (C) demonstrates the $R_g > L_p > d_{\text{gel}}$ limit, similar to the polyacrylamide system. In both schematics the electric field strength is illustrated by color gradient, with red being greatest electric field strength, and blue the smallest.

nanopore. Similarly, the short entropic trapping regime that replaces Ogston sieving when $R_g \approx d_{\text{gel}}$ does not offer many DNA-gel fiber interactions while the DNA molecule resides within the nanopore, and would not be conducive to slowing DNA translocation. For this reason, we focus on the reptation regime where $d_{\text{gel}} < R_g$ [59]. It is important to mention that there are several field-dependent subregimes that exist within the reptation regime. However, due to the unconventionally steep decay of the electric field near the mouth of the nanopore, we argue it would be nearly impossible to deconvolute them over the R_c length scale, and thus they can only be considered globally. Therefore we focus on the reptation regime under two conditions: (i) when the DNA molecule is flexible on the length scale of a gel pore ($d_{\text{gel}} > L_p$ where $L_p =$ 50 nm for dsDNA) and the polymer chain can easily deform when forced to invade the gel on the *trans* side; and (ii) when the DNA behaves as persistent (rod-like) segments in tight gel pores ($L_p > d_{\text{gel}}$) [59], and pushing the chain through the gel does not lead to conformational deformations. These two regimes are achieved through the use of agarose and polyacrylamide gels, respectively. Our experimental study will thus be able to explore these two different limits.

2 Materials and methods

2.1 Silicon nitride membranes

The silicon nitride (SiN_x) membranes utilized for our experiments were acquired from Norcada (Edmonton, Alberta), and are commercially available as transmission electron microscope windows. The membranes are composed of 10-nm thick low-stress SiN_x deposited on a 200-μm thick Si substrate by low pressure chemical vapor deposition. An anisotropic wet chemical etch is then used to produce a 50-μm × 50-μm freestanding membrane in the center of a 3-mm silicon support chip. Before depositing the gels on the membranes, the chips were cleaned in oxygen plasma for 1 min at 50 W.

2.2 Polymer gels

2.2.1 Agarose

1% w/v UltraPure agarose (Invitrogen, Carlsbad, CA) was dissolved in a heated 10 mM KCl solution (pH 8). The solution temperature was maintained at 60°C for ~45 min to ensure equilibrium had been reached. Any water that had evaporated was replaced with degassed DI water and the solution was briefly degassed again. A droplet (~2 μL) of the agarose solution was then pipetted onto the SiN_x membrane, with a silicon elastomer gasket offering containment and support to the liquid as it gelled. The resulting agarose gel thickness was estimated at ~1.5 mm. Note that the applied voltage was largely unaffected by the presence of the gel, as the resistance across it was measured to be < 1% of that of the agarose-interfaced nanopore system.

2.2.2 Polyacrylamide

Ultra-Pure Acrylamide, Bis, and TEMED were purchased from Fisher (Pittsburgh, PA). The concentration of acrylamide was 10%, and a Bis concentration of 0.8% was used. The gel also contained 380 mM Tris (pH 8.8), 10 mM KCl, 0.1% ammonium persulfate, and 0.1% TEMED. A droplet (~2 μL) was placed on the membrane, in the same manner as with the agarose. The gel was left for at least 3 h for polymerization to occur. Each gel was visually inspected and confirmed to be solid before chips were mounted in the fluidic cells. The measured resistance across the polyacrylamide gel constituted ~2% of the total resistance of the polyacrylamide-interfaced nanopore system.

2.3 Fluidic cell and mounting procedure

Once the gel solution had been placed on one side of the chip and allowed to solidify, the device was mounted in a polyether ether ketone fluidic cell, with custom-made silicon elastomer gaskets, ensuring a liquid-tight seal. The reservoir on either side of the chip was filled with ~250 μL of a 1 M KCl solution (pH 8), buffered with 10 mM HEPES. A pair of Ag/AgCl electrodes was used to apply the potential and measure the current, with the electrode on the *trans* side being grounded. All voltages are applied to the *cis* electrode with respect to ground. While a custom electronic circuit was employed during the nanopore fabrication process [60], an Axopatch 200B monitored the ionic current and supplied the voltage during the DNA experiments.

2.4 DNA experiments

Fifty, 100 bp, and 5 kbp dsDNA were purchased from Fermentas (Thermo Scientific, Pittsburgh, PA). The concentration of DNA employed for the experiments was ~6 ng/μL. Occasionally, the nanopores would become clogged and were cleared by reversing the voltage polarity. Custom LabView programs were used for data acquisition and analysis, and a National Instruments (Austin, TX) DAQ card (NI PXIe-6363) was used for data acquisition. The ionic current signal was acquired at 250 kHz and low-pass filtered at 100 kHz with a four-pole Bessel filter.

3 Results and discussion

We previously reported the fabrication of solid-state nanopores by controlled breakdown (CBD) of a thin dielectric membrane in an aqueous electrolyte environment, where an applied voltage produces an electric field strength > 0.5 V/nm and generates a sustained leakage current through an otherwise insulating membrane [40, 60]. In order to investigate the effects of a nanopore interfaced with a gel on DNA translocation kinetics, we deposit the gel medium on an intact SiN_x membrane, mount the gel-interfaced membrane in a fluidic cell and then fabricate a nanopore by CBD. We expect the dielectric breakdown process to be more likely to occur in a relatively fiber-free area at the gel/membrane interface, due to the greater availability of ions for charge injection [61] . In this case, a DNA strand exiting the nanopore will initially enter a solution-filled void before encountering gel fibers. The molecule will then begin the process of finding a route through the gel. Analysis of the results obtained from DNA translocation experiments are reported below.

3.1 Nanopore fabrication by CBD

Nanopore fabrication is carried out in 1 M KCl or 3.6 M LiCl (pH 8) with a constant applied voltage ranging from –7 to –9 V. During fabrication, a leakage current on the order of tens of nA is monitored. The charge transport mechanism through the dielectric membrane involves a form of trap-assisted electron tunneling [60] that results from charge injection at the electrolyte/membrane interface by specific redox reactions [61], the nature of which depends on the type of ions present within the electrostatic double layer surrounding the membrane. The fabrication of nanopores in SiN_x membranes interfaced with a gel reveals kinetics and a magnitude in leakage current similar to bare membranes under otherwise analogous conditions. Once the nanopore is successfully fabricated, its conductance is extracted from the slope of the I–V curve, which allows for the estimation of the effective diameter of the nanopore [61, 62].

Some nanopores formed in this manner display typical characteristics of bare CBD-fabricated nanopores, with well-defined breakdown events and nearly linear I–V curves. While these nanopores are stable and exhibit typical low-noise behavior, others behave anomalously, either clogging or rectifying strongly. Predictably, these anomalous nanopores produce current traces that are much noisier than those of ohmic nanopores. Another potential source of anomalous behavior—the delamination of the gel interface from the nanopore—was observed only once, resulting in no polyacrylamide-induced change in the translocation kinetics for 5 kbp dsDNA. Overall, the yield for formation of low-noise nanopores successfully interfaced with agarose or polyacrylamide gel was ~31% ($N = 16$) and ~42% ($N = 12$), respectively. Out of the low-noise gel-interfaced nanopores, ~60% ($N = 10$) produced a few hundred translocations at various voltages or over a thousand at a single voltage before irreversibly clogging. Since the CBD fabrication method produces nanopores capable of DNA detection with a yield of ~80% on bare membranes [60], we attribute this relatively low yield (~21%) to the presence of the gel.

3.2 Nanopores interfaced with agarose gel

We first consider the regime $R_g > d_{\text{gel}} > L_{p.}$ Here, a full-length DNA molecule must occupy several gel pores, with each gel pore being large enough to accommodate multiple flexible segments of DNA, as depicted in Fig. 1B. In order to satisfy these conditions, we use 1% agarose in 10 mM KCl, which results in an estimated d_{gel} of ~200–400 nm [63, 64], interfaced with a nanopore 3.6 nm in effective diameter, to which we introduce 5 kbp dsDNA (R_g ~470 nm and contour length of 1.7 μm). Note that this nanopore size is chosen to ensure single-file passage of dsDNA, while at the same time avoiding translocations in the friction-dominated regime [38, 40]. As shown in Fig. 2, agarose slows down some of the DNA molecules substantially, by as much as three orders of magnitude (from tens of μs to tens of ms), while

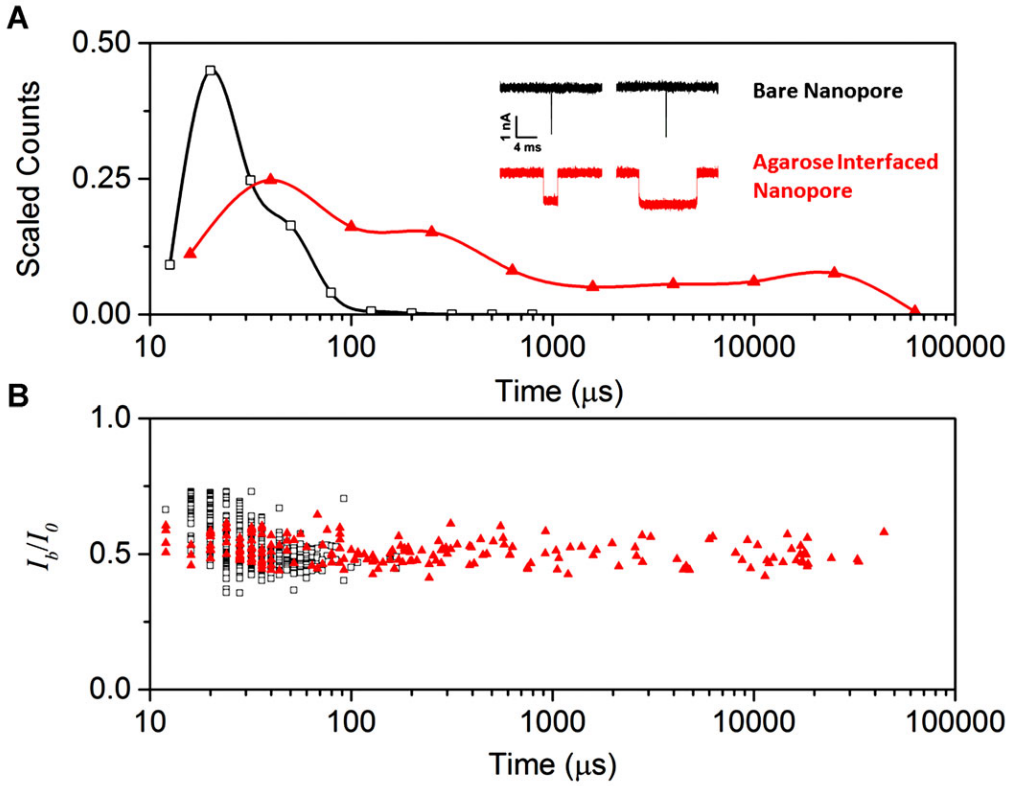


Figure 2. Comparison of 5 kbp dsDNA translocations through two similarly sized nanopores, one bare (4.1 nm) and the other with a *trans*-side agarose interface (3.6 nm). (A) Translocation time histograms reveal extended event durations for the agarose-interfaced nanopore versus the bare nanopore. Inset shows typical events for both cases. The events are 20 and 24 µs (left and right, respectively) for the bare nanopore, and 2800 and 10 440 µs for the agarose-interfaced nanopore (left and right, respectively). (B) Scatterplot of the current blockade versus translocation time shows similar blockage levels for the agarose-interfaced and bare nanopores. Each data point represents an individual DNA translocation event ($N_{\text{bare}} = 467$; $N_{\text{gel}} = 206$). All measurements were performed in 1 M KCl (pH 8) at –600 mV. Folded DNA translocations and partial translocations (collisions) were excluded for simplicity. Signals from molecules translocating more rapidly than ~20 µs will be increasingly truncated due to bandwidth limitations of the Axopatch 200B. This is observed in (B) and in Figs. 2 and 3, where the shallower blockages plotted for molecules translocating faster than ~20 µs are due to erroneous readings of the blockage depth (I_b). This does not affect our conclusions.

other DNA molecules pass through the gel with a dwell time similar to that observed in a bare nanopore of similar dimensions. The mean translocation time of the data plotted in Fig. 2A is calculated to be 35 ± 0.9 µs for the bare nanopore, while the agarose-interfaced nanopore has a mean translocation time nearly two orders of magnitude greater, at 2900 ± 470 µs. The wide distribution in the agarose-interfaced case is highlighted by the percentage of events longer than twice the mean; while roughly 3.5% of bare-nanopore events are longer than 70 µs, 15% of the agarose-interfaced nanopore events are longer than 5700 µs.

We attribute the overall increase in mean passage time to the DNA molecule interacting strongly with the gel on the *trans* side of the nanopore. There are two possible mechanisms for this interaction—electrical and steric. In terms of electrical (or charge-based) interactions, we might expect anion-binding centers in the agarose fibers [65] to interact with the highly negatively charged DNA molecules. However, the small number of sulfate groups (< 0.15%) combined with the screening effects of the high salt concentration in our agarose gel [66] ensure that this interaction will be minimal [67]. Instead, the principal mechanism for DNA-gel interactions should be steric in nature. During translocation through a bare nanopore, the leading end of a DNA molecule moves away from the nanopore while the rest of the molecule progresses through. The translocation kinetics in this case are usually dominated by the viscous drag of the portion of the molecule being pulled through the nanopore or by DNA-pore wall interactions [20, 68]. However, when the nanopore is interfaced with a gel, the biomolecule may be sterically hindered during passage, essentially resulting in the molecule crowding itself, where the portion of the DNA molecule that has not yet translocated through the nanopore cannot translocate until the fraction of the molecule on the *trans* side vacates the gel pore in the immediate vicinity of the nanopore.

While the shift in mean passage time is remarkable, the wide spread in the distribution suggests that some molecules are interacting with the gel differently than others, which can be undesirable for some applications and increases the

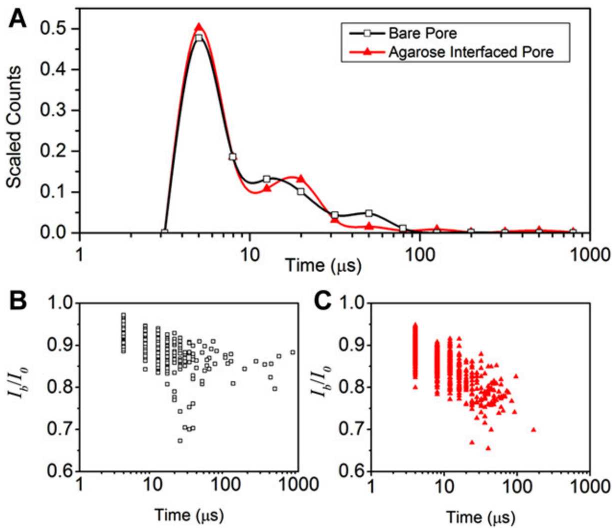


Figure 3. Comparison of 50 bp dsDNA translocations through two similarly sized nanopores, one bare (7.4 nm) and the other with a *trans*-side agarose interface (6.0 nm). (A) Translocation time histograms reveal nearly identical event durations for the agarose-interfaced nanopore versus the bare nanopore. Note that, due to limited bandwidth, it is very likely that many molecules are translocating undetected. The peak in the distribution should therefore not be regarded as the most probable passage time. (B) Scatterplot of the normalized current blockade versus translocation time shows very similar blockage levels for the agarose-interfaced and bare nanopores. Each data point represents an individual DNA translocation event ($N_{\text{bare}} = 726$; $N_{\text{gel}} = 821$). All measurements were performed in 3.6 M LiCl (pH 8) at –200 mV. Events corresponding to folded DNA translocations were excluded for clarity.

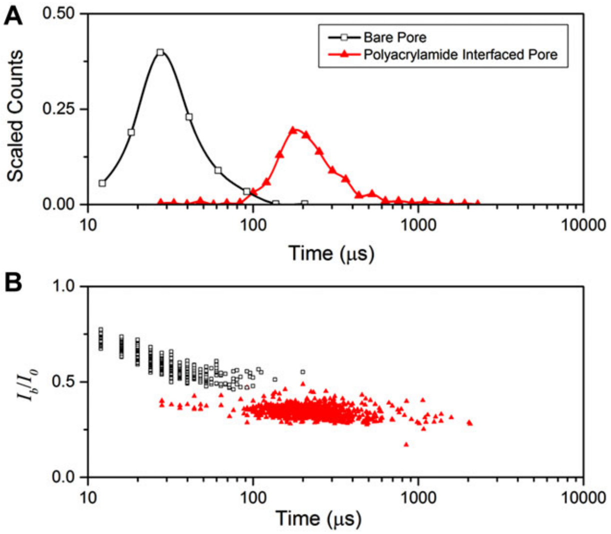


Figure 4. Comparison of 100 bp DNA translocations through two similarly sized nanopores, one bare (4.0 nm) and the other with a *trans*-side polyacrylamide interface (3.6 nm). (A) Translocation time histograms reveal extended event durations for the polyacrylamide-interfaced nanopore versus the bare nanopore. Note that, due to limited bandwidth, it is very likely that many molecules are translocating through the bare nanopore undetected. The peak in the distribution should therefore not be regarded as the most probable passage time for the case of the bare nanopore. (B) Scatterplot of the current blockade versus translocation time for both the polyacrylamide-interfaced and bare nanopores. Each data point represents an individual DNA translocation event ($N_{\text{bare}} = 518$; $N_{\text{gel}} = 847$). All measurements were performed in 3.6 M LiCl (pH 8) at –400 mV. Partial DNA translocations (collisions with the nanopore) were observed but excluded for clarity.

complexity of data analysis. We attribute this result to the stochastic behavior of DNA in agarose gels when subjected to high electric fields, specifically the propensity of DNA to undergo geometration (hooking) and herniation of loops (branching) around the agarose fibers [59]. These retardation mechanisms are randomly triggered by thermal fluctuations and/or sampling of various gel fibre configurations. Although numerous models can be used to predict DNA electrophoretic mobility within gels, they actually describe the mean effect of a large number of retardation events taking place over a considerable migration distance. In the context of slowing down the DNA translocation process in nanopores, however, each molecule can only be affected by a small number of retardation events, leaving little room to account for all of the different paths the DNA can take, which likely contributes to the spread of the distribution. In other words, differing translocation events arise from the various possible pathways through the gel. The large spread in translocation times should not be surprising, given that geometration and herniation, frequent under these conditions in the gel, introduce large variability in the DNA migration process [69–71]. For the same reasons, we do not anticipate all devices to behave identically. While we can easily fabricate a nanopore with controlled dimensions at the interface of the gel, we are unable to control the distribution of gel fibers near the nanopore. Therefore, while global trends pertaining to the increase in mean passage times and significant broadening of the distribution of translocation times are expected to persist, each device will show slightly different slowing characteristics as a result of variations in the gel (see Supporting Information Fig. 1).

To test the gel's slowing effects on a smaller molecule (with $R_g < d_{\text{gel}}$), we perform translocations with a 50 bp dsDNA fragment (rod-like with length of $\sim$17 nm) through a 6.0 nm nanopore interfaced with a 1% agarose gel. The resulting data agrees very well with a similar nanopore that was left bare (Fig. 3). Not accounting for the potentially large fraction of events passing through both the bare and the agarose-interfaced nanopores too rapidly to be detected, the mean observable translocation times are similar, at $\sim$10 μs. This is expected, as the 50 bp DNA only occupies a fraction of the void in the agarose, and therefore cannot both reside in the sensing region and interact with the gel fibers at the same time.

3.3 Nanopores interfaced with polyacrylamide gel

In response to the broad range of translocation times obtained with agarose-interfaced nanopores, we determined that much tighter gel pores were needed to provide more consistent steric hindrance during translocation events. Under these tight gel-pore conditions, herniation is unlikely to occur, so the molecule is forced to remain more or less linear, leading to fewer translocation conformations and reducing the spread of the translocation distribution. For this reason, we turn our attention toward the $L_p > d_{\mathrm{gel}}$ regime (as shown in Fig. 1C).

While entering this regime with agarose would be difficult, polyacrylamide, an equally well-characterized gel, is commonly employed in gel electrophoresis and possesses smaller gel pores (d_{gel} ~5–100 nm, with average gel pore size likely to be ~10 nm [59]). This should increase the likelihood of DNA-gel interactions while reducing the issue of herniation. Thus we expect that DNA will translocate into a more confined environment in polyacrylamide than it does in agarose, leading to more predictable translocation characteristics. This expectation is confirmed by the analysis of 100 bp DNA (rod-like with a length of ~34 nm) translocations through a nanopore interfaced with polyacrylamide. As shown in Fig. 4, a 3.6-nm nanopore interfaced with a polyacrylamide gel strongly interacts with the rod-like DNA, essentially slowing each 100 bp molecule. The mean translocation time for the polyacrylamide-interfaced nanopore is measured to be 248 ± 6 μs, nearly an order of magnitude slower than the 33 ± 1 μs mean passage time through a similarly sized bare nanopore. This translates into ~0.4 bp/μs, a step toward achieving the necessary translocation speed for DNA sequence analysis. Importantly, the spread in translocation times is, in this case, spanning only ~170 μs (as calculated from the Full Width Half Maximum, FWHM). The preservation of the distribution spread is highlighted by the number of translocations at twice the mean; for the bare nanopore, ~4.8% of the events are slower than 66 μs, while for the polyacrylamide-interfaced nanopore, ~5.5% of the events are slower than 496 μs. However, as previously mentioned, the precise size of the gel pores in the vicinity of the nanopore is not easily controlled, and it is expected that the distribution of gel pore size encountered by the DNA as it moves out of the nanopore and through the gel, will influence the details of the translocation kinetics. In fact, Supporting Information Fig. 2 reveals that not all polyacrylamide-interfaced nanopore devices will behave identically. While the most probable passage time is still shifted by over an order of magnitude, the histogram of translocation times contains a characteristic shoulder that widens the distribution considerably. Upon closer inspection, we find that the long duration translocation events occur exclusively when the molecule translocates immediately after the previous one (Supporting Information Fig. 3). The observed correlation between translocation time and the time elapsed since the previous translocation is consistent with the idea that the translocating molecule can interact with other molecules within the *trans* side gel. This is not surprising, as the purpose of utilizing a *trans* side gel is to restrict the mobility of DNA during and after translocation.

The significant structural differences between agarose and polyacrylamide are further emphasized by efforts to translocate the much longer 5 kbp dsDNA molecules. As described earlier, an agarose interface slowed the mean translocation time of 5 kbp fragments; however, the polyacrylamide interface slows the 5 kbp fragments to the extent that very few are even able to complete full translocation through the nanopore, and instead remain within the nanopore for several minutes until the voltage is altered to clear the nanopore.

4 Concluding remarks

We showed that DNA translocations through a solid-state nanopore can be slowed down by interfacing the membrane on the *trans* side with a well characterized and inexpensive gel medium—agarose or polyacrylamide—prior to fabricating a nanopore by CBD. A nanopore interfaced with agarose in this way produced a mean translocation time for 5 kbp dsDNA nearly two orders of magnitude slower than a similarly sized bare nanopore operated under the same conditions. However, the distribution of event lengths observed with the agarose-interfaced nanopore spanned over three orders of magnitude. This large variance in the translocation times suggests that the agarose gel does not offer the continuous friction needed in order to slow down the molecules uniformly, and that there are simply too many routes for the DNA to take once inside the gel. Consequently, polyacrylamide-interfaced nanopores were used in order to exploit their tighter gel pores. These presented an order-of-magnitude shift in the mean DNA translocation time with nearly every individual 100 bp dsDNA molecule being slowed, permitting observation of the entire distribution of translocation times despite the short length of the fragments and the relatively high voltages used. We attribute the slowing efficiency of the polyacrylamide to a higher probability of interactions between DNA and polyacrylamide fibers, which presents a more stable network of tighter gel pores compared to agarose. Additionally, intermolecular crowding on the *trans* side, a rather unlikely interaction for typical translocation experiments, was found to contribute an additional retardation mechanism for some nanopores. However, when inter-DNA crowding was present on the *trans* side, it came with an additional broadening mechanism.

In summary, the use of a gel medium interfaced on the *trans* side of a solid-state nanopore significantly slows the translocation speed of DNA without affecting the capture process from the *cis* side. The use of commonly employed agarose and polyacrylamide as gel media ensures that this approach is both economical and accessible. In future work, the gel medium could be used to recapture previously translocated biomolecules by reversing the polarity of the voltage for repeated analysis, or preloaded with molecules as a means to rapid nanopore-based analysis on minute quantities of biological sample.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Foundation for Innovation. K. Briggs and M. Waugh acknowledge the financial support provided by NSERC, and OGS, *respectively, for Postgraduate Fellowships. The authors would like to thank Dr. James Harden for kindly granting the use of his polyacrylamide reagents.*

The authors have declared no conflict of interest.

5 References

[1] Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V, Ramsey, J. M., Riehn, R., Soni, G. V, Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J. A., *Nat. Biotechnol.* 2008, *26*, 1146.

[2] Deamer, D. W., Branton, D., *Acc. Chem. Res.* 2002, *35*, 817.

[3] Zwolak, M., Di Ventra, M., *Rev. Mod. Phys.* 2008, *80*, 141.

[4] Fologea, D., Uplinger, J., Thomas, B., McNabb, D. S., Li, J., *Nano Lett.* 2005, *5*, 1734.

[5] Venkatesan, B. M., Bashir, R., *Nat. Nanotechnol.* 2011, *6*, 615.

[6] Deamer, D., *Annu. Rev. Biophys.* 2010, *39*, 79.

[7] Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K., Akeson, M., *Nat. Biotechnol.* 2012, *30*, 344.

[8] Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., Pavlenok, M., Niederweis, M., Gundlach, J. H., *Nat. Biotechnol.* 2012, *30*, 349.

[9] Zhang, X., Wang, Y., Fricke, B. L., Gu, L., *ACS Nano* 2014, *8*, 3444.

[10] Singer, A., Rapireddy, S., Ly, D. H., Meller, A., *Nano Lett.* 2012, *12*, 1722.

[11] Meller, A., Nivon, L., Brandin, E., Golovchenko, J., Branton, D., *Proc. Natl. Acad. Sci. USA* 2000, *97*, 1079.

[12] Fologea, D., Uplinger, J., Thomas, B., Mcnabb, D. S., Li, J., *Nano Lett.* 2005, *5*, 1734.

[13] Kawano, R., Schibel, A. E. P., Cauley, C., White, H. S., *Langmuir* 2009, *25*, 1233.

[14] Luan, B., Wang, D., Zhou, R., Harrer, S., Peng, H., Stolovitzky, G., *Nanotechnology* 2012, *23*, 455102.

[15] Kowalczyk, S. W., Wells, D. B., Aksimentiev, A., Dekker, C., *Nano Lett.* 2012, *12*, 1038.

[16] Smeets, R. M. M., Keyser, U. F., Krapf, D., Wu, M.-Y., Dekker, N. H., Dekker, C., *Nano Lett.* 2006, *6*, 89.

[17] Hatlo, M. M., Panja, D., van Roij, R., *Phys. Rev. Lett.* 2011, *107*, 068101.

[18] Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y., Meller, A., *Nat. Nanotechnol.* 2010, *5*, 160.

[19] Zhang, H., Zhao, Q., Tang, Z., Liu, S., Li, Q., Fan, Z., Yang, F., You, L., Li, X., Zhang, J., Yu, D., *Small* 2013, *9*, 4112.

[20] Lu, B., Hoogerheide, D. P., Zhao, Q., Zhang, H., Tang, Z., Yu, D., Golovchenko, J., *Nano Lett.* 2013, *13*, 3048.

[21] Mitchell, N., Howorka, S., *Angew. Chemie* 2008, *120*, 5647.

[22] Lieberman, K. R., Cherf, G. M., Doody, M. J., Olasagasti, F., Kolodji, Y., Akeson, M., *J. Am. Chem. Soc.* 2010, *132*, 17961.

[23] Chu, J., González-López, M., Cockroft, S. L., Amorin, M., Ghadiri, M. R., *Angew. Chem. Int. Ed. Engl.* 2010, *49*, 10106.

[24] Cockroft, S. L., Chu, J., Amorin, M., Ghadiri, M. R., *J. Am. Chem. Soc.* 2008, *130*, 818.

[25] Hornblower, B., Coombs, A., Whitaker, R. D., Kolomeisky, A., Picone, S. J., Meller, A., Akeson, M., *Nat. Methods* 2007, *4*, 315.

[26] Derrington, I. M., Butler, T. Z., Collins, M. D., Manrao, E., Pavlenok, M., Niederweis, M., Gundlach, J. H., *Proc. Natl. Acad. Sci. USA* 2010, *107*, 16060.

[27] Mathé, J., Visram, H., Viasnoff, V., Rabin, Y., Meller, A., *Biophys. J.* 2004, *87*, 3205.

[28] Dudko, O. K., Mathé, J., Szabo, A., Meller, A., Hummer, G., *Biophys. J.* 2007, *92*, 4188.

[29] Vercoutere, W., Winters-Hilt, S., Olsen, H., Deamer, D., Haussler, D., Akeson, M., *Nat. Biotechnol.* 2001, *19*, 248.

[30] Olasagasti, F., Lieberman, K. R., Benner, S., Cherf, G. M., Dahl, J. M., Deamer, D. W., Akeson, M., *Nat. Nanotechnol.* 2010, *5*, 798.

[31] Wiggin, M., Tropini, C., Tabard-Cossa, V., Jetha, N. N., Marziali, A., *Biophys. J.* 2008, *95*, 5317.

[32] Jetha, N. N., Feehan, C., Wiggin, M., Tabard-Cossa, V., Marziali, A., *Biophys. J.* 2011, *100*, 2974.

[33] Nakane, J., Wiggin, M., Marziali, A., *Biophys. J.* 2004, *87*, 615.

[34] Van den Hout, M., Vilfan, I. D., Hage, S., Dekker, N. H., *Nano Lett.* 2010, *10*, 701.

[35] Keyser, U. F., Koeleman, B. N., van Dorp, S., Krapf, D., Smeets, R. M. M., Lemay, S. G., Dekker, N. H., Dekker, C., *Nat. Phys.* 2006, *2*, 473.

[36] Peng, H., Ling, X. S., *Nanotechnology* 2009, *20*, 185101.

[37] Hyun, C., Kaur, H., Rollings, R., Xiao, M., Li, J., *ACS Nano* 2013, *7*, 5892.

[38] Wanunu, M., Sutin, J., McNally, B., Chow, A., Meller, A., *Biophys. J.* 2008, *95*, 4716.

[39] Kurz, V., Nelson, E. M., Shim, J., Timp, G., *ACS Nano* 2013, *7*, 4057.

[40] Briggs, K., Kwok, H., Tabard-Cossa, V., *Small* 2014, *10*, 2077.

[41] Di Fiori, N., Squires, A., Bar, D., Gilboa, T., Moustakas, T. D., Meller, A., *Nat. Nanotechnol.* 2013, *8*, 946.

[42] Krishnakumar, P., Gyarfas, B., Song, W., Sen, S., Zhang, P., Krstić, P., Lindsay, S., *ACS Nano* 2013, *7*, 10319.

[43] Yusko, E. C., Johnson, J. M., Majd, S., Prangkio, P., Rollings, R. C., Li, J., Yang, J., Mayer, M., *Nat. Nanotechnol.* 2011, *6*, 253.

[44] Anderson, B. N., Muthukumar, M., Meller, A., *ACS Nano* 2013, *7*, 1408.

[45] He, Y., Tsutsui, M., Scheicher, R. H., Bai, F., Taniguchi, M., Kawai, T., *ACS Nano* 2013, *7*, 538.

[46] Belkin, M., Maffeo, C., Wells, D. B., Aksimentiev, A., *ACS Nano* 2013, *7*, 6816.

[47] Luan, B., Peng, H., Polonsky, S., Rossnagel, S., Stolovitzky, G., Martyna, G., *Phys. Rev. Lett.* 2010, *104*, 238103.

[48] Polonsky, S., Rossnagel, S., Stolovitzky, G., *Appl. Phys. Lett.* 2007, *91*, 153103.

[49] Ai, Y., Liu, J., Zhang, B., Qian, S., *Anal. Chem.* 2010, *82*, 8217.

[50] Paik, K., Liu, Y., Tabard-cossa, V., Waugh, M. J., Huber, D. E., Provine, J., Howe, R. T., Dutton, R. W., Davis, R. W., 2012, 6767.

[51] Liu, Y., Huber, D. E., Tabard-Cossa, V., Dutton, R. W., *Appl. Phys. Lett.* 2010, *97*, 143109.

[52] He, Y., Tsutsui, M., Fan, C., Taniguchi, M., Kawai, T., *ACS Nano* 2011, *5*, 5509.

[53] Yen, P., Wang, C., Hwang, G.-J., Chou, Y. C., *Rev. Sci. Instrum.* 2012, *83*, 034301.

[54] Tsutsui, M., He, Y., Furuhashi, M., Rahong, S., Taniguchi, M., Kawai, T., *Sci. Rep.* 2012, *2*, 394.

[55] Tsutsui, M., Rahong, S., Iizumi, Y., Okazaki, T., Taniguchi, M., Kawai, T., *Sci. Rep.* 2011, *1*, 46.

[56] Squires, A. H., Hersey, J. S., Grinstaff, M. W., Meller, A., *J. Am. Chem. Soc.* 2013, *135*, 16304.

[57] Kang, X., Cheley, S., Rice-Ficht, A. C., Bayley, H., *J. Am. Chem. Soc.* 2007, *129*, 4701.

[58] Ando, G., Hyun, C., Li, J., Mitsui, T., *ACS Nano* 2012, *6*, 10090.

[59] Viovy, J., *Rev. Mod. Phys.* 2000, *72*, 813.

[60] Kwok, H., Briggs, K., Tabard-Cossa, V., *PLoS One* 2014, *9*, e92880.

[61] Briggs, K., Charon, M., Kwok, H., Le, T., Chahal, S., Bustamante, J., Waugh, M., Tabard-Cossa, V., *Nanotechnology* 2015, *26*, 084004.

[62] Kowalczyk, S. W., Grosberg, A. Y., Rabin, Y., Dekker, C., *Nanotechnology* 2011, *22*, 315101.

[63] Holmes, D. L., Stellwagen, N. C., *Electrophoresis* 1990, *11*, 5.

[64] Maaloum, M., Pernodet, N., Tinland, B., *Electrophoresis* 1998, *19*, 1606.

[65] Stellwagen, N. C., *Electrophoresis* 2009, *30*, S188.

[66] Fatin-Rouge, N., Milon, A., Buffle, J., Goulet, R. R., Tessier, A., *J. Phys. Chem. B* 2003, *107*, 12126.

[67] Piculell, L., Nilsson, S., *J. Phys. Chem.* 1989, *93*, 5602.

[68] Wanunu, M., Sutin, J., Meller, A., *Nano Lett.* 2009, *9*, 3498.

[69] Starchev, K., Sturm, J., Weill, G., Brogren, C.-H., *J. Phys. Chem. B* 1997, *101*, 5659.

[70] Fatin-Rouge, N., Starchev, K., Buffle, J., *Biophys. J.* 2004, *86*, 2710.

[71] Saxton, M. J., *Biophys. J.* 1994, *66*, 394.

D

APPENDIX: Free energy of a polymer in slit-like confinement from the Odijk regime to the bulk

Jason S. Leith, Albert Kamanzi, **David Sean**, Daniel Berard, Andrew C. Guthrie, Christopher M.J. McFaul, Gary W. Slater, Hendrick W. de Haan and Sabrina R. Leslie

Published in *Macromolecules*
Jason S. Leith, Albert Kamanzi, David Sean, Daniel Berard, Andrew C. Guthrie, Christopher M.J. McFaul, Gary W. Slater, Hendrick W. de Haan and Sabrina R. Leslie: Free energy of a polymer in slit-like confinement from the Odijk regime to the bulk. Macromolecules. 2016, **49**. 9266–9271.

http://dx.doi.org/10.1021/acs.macromol.6b01805

Macromolecules
Article
pubs.acs.org/Macromolecules

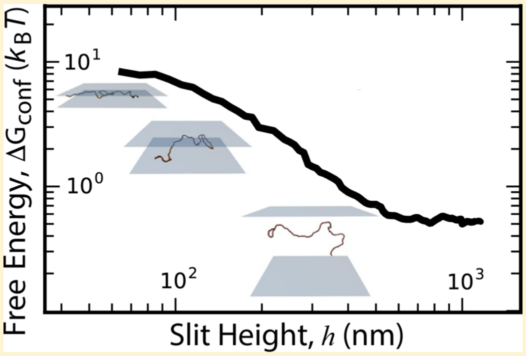
Free Energy of a Polymer in Slit-like Confinement from the Odijk Regime to the Bulk

Jason S. Leith,[†] Albert Kamanzi,[†] David Sean,[‡] Daniel Berard,[†] Andrew C. Guthrie,[†] Christopher M. J. McFaul,[†] Gary W. Slater,[‡] Hendrick W. de Haan,[*,§] and Sabrina R. Leslie[*,†]

[†]McGill University, Montreal, Quebec, Canada H3A 2T8
[‡]University of Ottawa, Ottawa, Ontario, Canada K1N 6N5
[§]Institute of Technology, University of Ontario, Oshawa, Ontario, Canada L1H 7K4

Supporting Information

ABSTRACT: We directly measure the free energy of confinement for semiflexible polymers from the nanoscale to bulk regimes in slit-like confinement. We use convex lens-induced confinement (CLiC) microscopy of DNA to directly count molecules at equilibrium in a single chamber of smoothly increasing height. Our data, acquired across a continuum of confinement regimes, provide a bridge with which to connect scaling theories established for qualitatively different regimes. We present new experimental data and simulations that connect the Odijk theory describing sub-persistence-length confinement, the interpolation model by Chen and Sullivan extending Odijk to moderate confinement, and the Casassa theory describing the transition from moderate confinement to bulk. Further, this work establishes a robust, quantitative platform for understanding and manipulating biopolymers at the nanoscale, with key applications and insights toward emerging genomic analysis tools.



■ INTRODUCTION

A detailed understanding of the behavior of polymers under confinement has critical applications in the fields of nanotechnology, genomics, biophysics, and materials science. Particularly, direct manipulation and separation of biopolymers have contributed to the development of advanced nucleic acid analytical technologies[1,2] which integrate nanofluidics techniques such as extensional flow,[3] physical confinement,[4] and molecular combing.[5]

The key to these efforts has been technical innovations in nanofluidic confining devices. Among them are nanochannels,[6] nanopillar arrays,[7] nanoslits,[8] and staircase-like devices.[9] Experimental studies using these systems have generally employed long ($\gg$ hundreds of persistence lengths) polymers. The high free energy cost of confining such long polymers under the strongest-confinement regime, the Odijk regime,[10] makes it challenging to make equilibrium measurements of this free energy without using indirect techniques.[11] Direct measurements under this strong confinement require advances in populating highly confined geometries, which can be facilitated by using shorter polymers. Further, experiments with short polymers expand on prior verification of theoretical and simulation-based studies.[12]

Historically, slit confinement of polymer molecules has been divided into three regimes: strong confinement, or the Odijk regime,[10] moderate confinement, and weak confinement or the bulk.[13] These regimes are defined by comparing the polymer's unconfined radius of gyration, R_g, and Kuhn length, L_K, to the confining height h. The least-confined regime is the bulk regime, $h \gg 2R_g$, in which the polymer generally is not deformed from its free-solution conformation, except possibly when it comes close to a wall.

In the Odijk regime, $h \ll L_K$, and the polymer is stiff for distances on the order of the slit height. As a consequence, the polymer is unable to form random coils like it does in free solution, and its conformations are affected by *deflecting* stiff segments off of the walls (Figure 1c, left). Odijk[10,14] predicted that for cylindrical tube-like confinement, as opposed to slit-like confinement, the free energy of confinement, G_{conf}, in this regime scales as

$$\frac{G_{\mathrm{conf}}(h)}{k_BT} \propto \frac{L_c}{L_K}\left(\frac{L_K}{h}\right)^{2/3} \tag{1}$$

where h is the diameter of the cylinder, k_B is the Boltzmann constant, T is the temperature, and L_c is the contour length of the polymer. The free energy of a polymer in a rectangular tube was later shown to be the sum of two terms, each identical to eq 1, except with h replaced in either term by the length of one

Received: August 19, 2016
Revised: November 15, 2016
Published: December 2, 2016

DOI: 10.1021/acs.macromol.6b01805

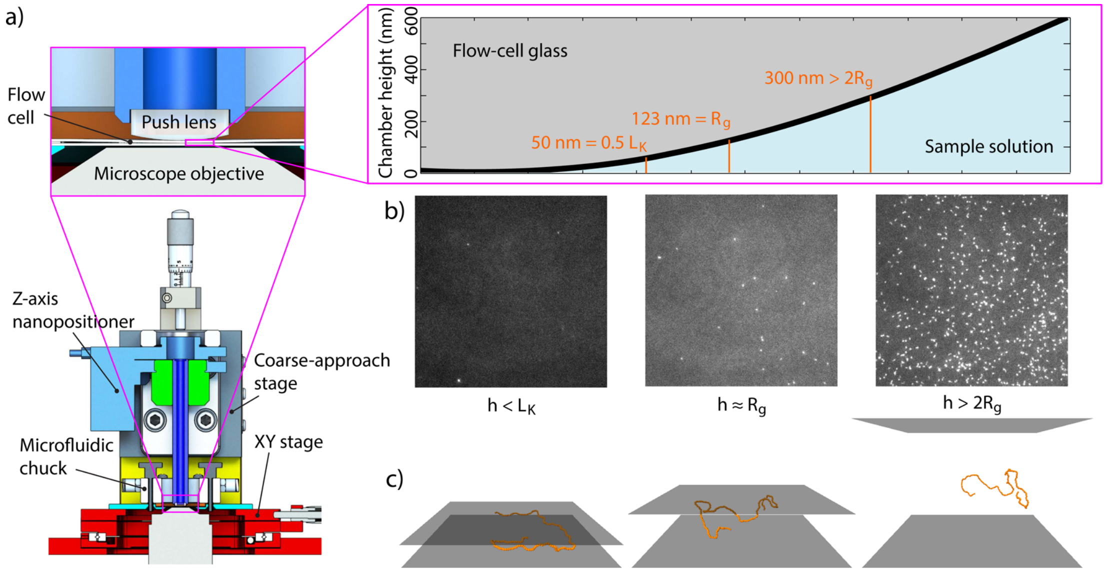


Figure 1. (a) Schematic of the CLiC imaging device: includes an imaging flow cell, a "push lens" for forming the chamber, and a chuck with sample inlets for flowing in samples. It is mounted above the imaging objective. (b) 91 × 91 nm^2 micrographs of pUC19 dsDNA molecules with unconfined radius of gyration, R_g = 123 nm, located at different points in the CLiC chamber. Left panel shows a chamber region with height $h < L_K$ (the Odijk regime); center panel shows a region centered roughly on $h = R_g$ (the moderate-confinement regime); right panel shows a region with $2R_g < h < 3R_g$ (transition between moderate-confinement and bulk regimes). (c) Renderings of different polymer conformations in different regimes obtained by simulation. The three orange vertical lines in the chamber profile [(a), upper right] indicate heights at which the three simulated polymers respectively are confined.

of the two sides of the rectangle[15] and with prefactors to the scaling relationship determined.

In the moderate-confinement regime ($L_K \ll h \ll 2R_g$), the polymer has enough room to bend back on itself in all three dimensions. Here, free energy is predicted[16,17] to scale with slit height according to

$$\frac{G_{\mathrm{conf}}(h)}{k_B T} \propto \left(\frac{R_g}{h}\right)^2 \tag{2}$$

where we have assumed an ideal polymer since the effects of self-avoidance are weak for short chains.[18] Note that previous studies indicate that in this moderate-confinement regime the free energy for finite chains will scale with the steric exclusion radius instead of the radius of gyration.[19] However, our concern here is the scaling with h as we do not vary the polymer size but rather the confinement dimensions. This regime has a lower bound given by the structural scale L_K and an upper bound of $2R_g$, which is dependent on the length of the polymer. For semiflexible chains that are short, the lack of sufficient separation between these length scales can lead to a regime that is too narrow to clearly give rise to the pure scaling (as opposed to the transitions).

In describing the behavior of polymers from extreme nanoconfinement to bulk regimes, we must therefore study two transition regions: (i) h between Odijk ($h \ll L_K$) and moderate confinement ($h \gg L_K$) and (ii) h between moderate confinement ($h < 2R_g$) and bulk ($h \gg 2R_g$). The first transition (which we will call the "CS transition") was studied by Chen and Sullivan,[12] who calculated the transition from moderate-confinement to Odijk regimes for an infinitely long polymer and proposed an empirical equation:

$$\frac{G_{\mathrm{conf}}(h)}{k_B T} = \frac{\pi^2}{6}\frac{L_c}{L_K}\frac{\left(\frac{L_K}{h}\right)^2}{\left[C_1\left(\frac{L_K}{h}\right)^2 + C_2\left(\frac{L_K}{h}\right) + 1\right]^{2/3}} \tag{3}$$

where C_1 = 1.2865 and C_2 = 0.992.[12] Chen and Sullivan's result was later confirmed using a Monte Carlo method for long polymers.[20]

For the second transition (which we will call the "Casassa transition"), a prediction for the conformational cost of an ideal polymer between impenetrable plates was given by Casassa.[13] In free energy terms, it can be expressed as

$$\frac{G_{\mathrm{conf}}(h)}{k_B T} = -\ln\left[\frac{8}{\pi^2}\sum_{m=0}^{\infty}\frac{\exp\left(-\frac{(2m+1)^2\pi^2}{4}\left(\frac{2R_g}{h}\right)^2\right)}{(2m+1)^2}\right] \tag{4}$$

which agrees with moderate confinement scaling of eq 2 in the $h/R_g \to 0$ limit. At large heights, a polymer will experience confinement only when it diffuses near the walls. This can be viewed as a narrow exclusion layer near the surfaces, which for hard spheres would equal their radii. As $h \to \infty$, this is a diminishing contribution which scales as $G_{\mathrm{conf}}(h) \propto k_B T(R_g/h)$.

An outstanding challenge in the field of polymer physics has been to experimentally resolve the transitions between all of the above confinement regimes. Experimental measurements of spatially resolved polymers have observed a CS transition at $h \approx L_K$ in the scaling of polymers' in-plane radius of gyration, $R_{\parallel}(L_c)$, first in channel-like confinement[21] and later in slit-like confinement.[22] Theoretical studies assessing whether this observed transition in the scaling of $R_{\parallel}$ corresponds to the predicted transition in free energy scaling have given less clear results.[23−25] One study[24] even suggests that the CS free energy transition may not exist in slit-like confinement.

Testing this theory requires loading and observing the same polymer under a continuum of applied confinement. It is hard, however, to populate nanoscale imaging volumes where molecules are excluded by a large potential, especially while keeping polymers intact.

In the present experiments, we overcome this challenge by using dsDNA fragments of contour length 1047 nm ≈ $10L_K$. Polymers of this size sufficiently populate regions of low height for adequate statistics in more strongly confined regimes, while maintaining a sufficiently low bulk concentration for molecules to avoid interacting with each other.

In this article, we experimentally measure the transition between the bulk and moderate-confinement regimes and the crossover into the Odijk regime using a single-molecule imaging and manipulation platform called convex lens-induced confinement (CLiC).[26,27] CLiC microscopy allows for a range of confinement heights from the nanoscale to the tens of micrometers to be probed simultaneously in one device. It loads molecules quickly, yet gently, into nanoscale environments, overcoming limitations of traditional side-loading nanofluidics.[1]

■ EXPERIMENTAL SECTION

The CLiC system (Figure 1a) enables direct counting of individual DNA molecules in a chamber of tunable height profile and hence tunable particle concentration profile. The molecules are loaded into a flow cell, initially between 10 and 30 μm thick, which is compressed to make a curved chamber height profile of shallow curvature (see Supporting Information). The height of the chamber can be essentially zero at the center (to within the surface roughness of the flow cell glass, 0.3−0.5 nm) and several micrometers at the edges. This allows our molecules of L_c = 1047 nm having a free-solution size of R_g = 123 nm (using ideal worm-like chain theory, e.g., as in ref 28) to quickly equilibrate among all regimes of confinement. CLiC microscopy produces a continuum of confinement heights without the need for micro- or nanofabrication, unlike devices with pillars[7] or staircases.[9] An earlier work of ours employed an additional mechanism, electrophoresis, to populate the most confined regions of the chamber for direct observation[29] and made measurements of DNA concentration away from equilibrium; the present work relies on no such mechanism and as such allows us to measure the free energy solely of *confinement*.

To study DNA that is both under strong confinement and in equilibrium with DNA under weak confinement, we used linearized pUC19 plasmids labeled with the fluorescent stain YOYO-1. The YOYO-1-labeled DNA, along with dilute, spectrally disjunct Cy5 dye, used for a fluorescence intensity signal proportional to local chamber height, was flowed into a CLiC chamber and allowed to equilibrate for over an hour. This period was sufficient to establish equilibrium concentration profiles which were measured reproducibly as a function of chamber height (see Supporting Information). Data collection involved a sequence of raster scans over the chamber of the Cy5 fluorescence (Figure S1), labeled-DNA fluorescence (Figure S2), and interferometry (Figure S1), in that order.

To obtain accurate measurements of the chamber height, h, we mapped the chamber geometry by fitting Cy5 fluorescence intensity to a polynomial function subject to constraints imposed by interferometry of the chamber, building on our methods in ref 27 (see Supporting Information). We then binned the height map into areas of approximately equal height-ranges, forming annuli about the center of the chamber (Figure S2).

The DNA concentration profile as a function of local height was obtained using an ImageJ particle tracking plugin originally developed by Sbalzarini and Koumoutsakos,[30] which we modified for this work (see Supporting Information). The trajectories were manually inspected to eliminate false particles and to include any missed particles. Particles were tracked for a duration of 2.5 s (50 frames at 50 ms exposure).

For a given bulk concentration, there is a limited range of heights where data can be taken. A practical lower limit is imposed at very small heights by the scarcity of DNA (first panel of Figure 1b), and an upper limit by particles being too dense to be reliably identified and tracked, or by exceeding the optical system's depth of field, resulting in particles diffusing in and out of focus. Since the CS and Casassa transitions for our experiments' R_g, L_K, and L_c span a large range of heights,[12,20] we performed experiments using several bulk concentrations in order to count particles at each height (Figure 2, top). The free energy profile was constructed piecewise using these data sets taken with overlapping observable height ranges and different bulk concentrations.

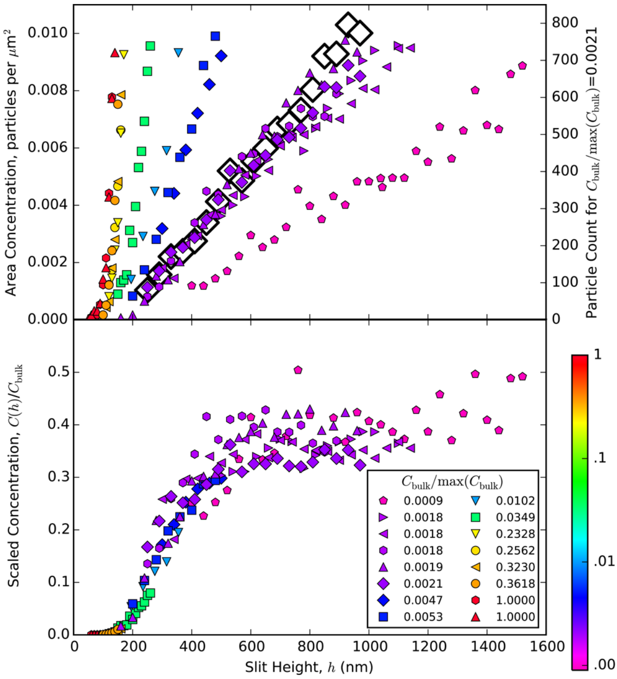


Figure 2. (top) Colored markers, left-hand axis: area concentration of particles in bins of height h. Open markers, right-hand axis: total number of particles counted for a selected data set. The bulk concentration, C_{bulk}, for the 16 different experiments, was varied to investigate a wide range of confinement heights. C_{bulk} reported as fraction of max(C_{bulk}), 12.9 nM, determined for the highest-concentration experiments, described in the Supporting Information. (bottom) Particle concentration $C(h)$ normalized by C_{bulk}.

We divided our particle counts by the volume of the relevant annulus to find the resulting DNA concentration (Figure 2, bottom), where the local concentration $C(h)$ is normalized by the bulk concentration C_{bulk}. Establishing the true C_{bulk} of a data set is complicated by lack of an *in situ* measurement and is achieved with a best fit described in the Supporting Information. The true C_{bulk} is lower than the concentration of the injected sample due to a small fraction of molecules adhering to device surfaces and incomplete replacement of the wetting buffer (further details in the Supporting Information).

■ RESULTS AND DISCUSSION

Figure 3 shows the confinement potential measured via CLiC as a function of confinement height h as well as the predicted values of G_{conf} from the Chen−Sullivan (eq 3) and Casassa (eq 4) theories. The confinement energy G_{conf} is calculated using

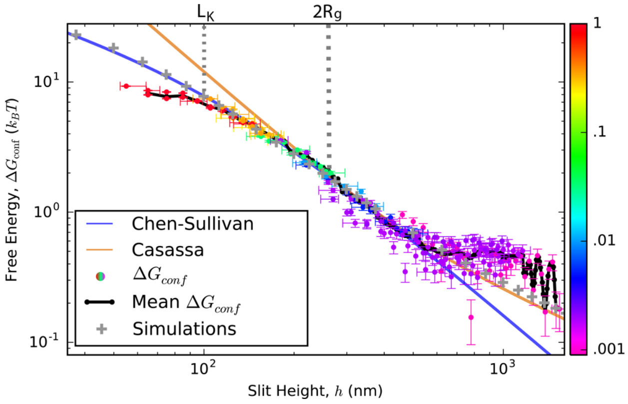


Figure 3. Confinement free energy for experiments, simulations, and CS and Casassa theories as a function of the slit height h. The Casassa line has been shifted down to meet the CS line to give a smooth crossover. Filled markers are experimental data coded as in Figure 2. Absolute concentrations were determined by a one-parameter fit of the combined data sets to a combined theory curve defined by the Chen–Sullivan (CS) curve at heights lower than the height at which the CS and Casassa curves coincide and by the Casassa curve at greater heights. Vertical error bars indicate the ranges of ΔG_{conf} corresponding to one standard deviation in the fit parameter (Supporting Information). Horizontal error bars indicate the range from the 25th to the 75th percentile of heights implied by the Cy5 fluorescence intensity within an annular height bin as defined by the polynomial chamber fit (see Supporting Information). For clarity, horizontal error bars are shown only for the lowest, highest, and central (rounded downward) height bins for each data set. Gray crosses are simulation data. Vertical dotted lines mark regime boundaries L_K and $2R_g$.

the DNA concentrations and the Boltzmann distribution expression

$$\frac{G_{\text{conf}}}{k_BT} = -\ln\left(\frac{C(h)}{C_{\text{bulk}}}\right) \tag{5}$$

Using the CLiC setup, data have been acquired down to $h < L_K$. As can be seen, this is well into the crossover regime where the scaling transitions from $G_{\text{conf}} \sim h^{-2}$ in the moderate-confinement regime to $G_{\text{conf}} \sim h^{-2/3}$ in the Odijk regime. The experimental data follow the smooth, gradual change predicted by Chen–Sullivan, albeit with a discrepancy at very low h, representing a direct observation of the CS transition for slit-like confinement.

We complement these experimental data with Monte Carlo simulations. An attempt is made to build a persistent random walk of ≈10 Kuhn segments (see Supporting Information for details). Any random walk that crosses the walls is rejected, and the free energy can be calculated from the ratio between the number of allowed conformations to the number of attempts. The simulation results for G_{conf} are shown in Figure 3 (gray crosses). The results agree with the CS transition from Odijk to moderate confinement and then smoothly follow the Casassa transition into the bulk scaling of h^{-1}.

Note that eq 4 is derived for a freely jointed chain while these simulations are performed with a semiflexible chain which could modify the prefactor (see Supporting Information for details). Further, the Casassa result is derived for infinitely long chains, and previous work has shown that finite chains can exhibit a different prefactor.[19] In any case, the data in Figures 3 and 4 demonstrate that the difference is quite negligible, and thus it is ignored here. The agreement between the Chen–Sullivan theory, the CLiC data, and the simulations indicates that although eq 3 is derived for infinitely long chains, it is also remarkably accurate for chains such as pUC19 which comprise only 10 Kuhn lengths.

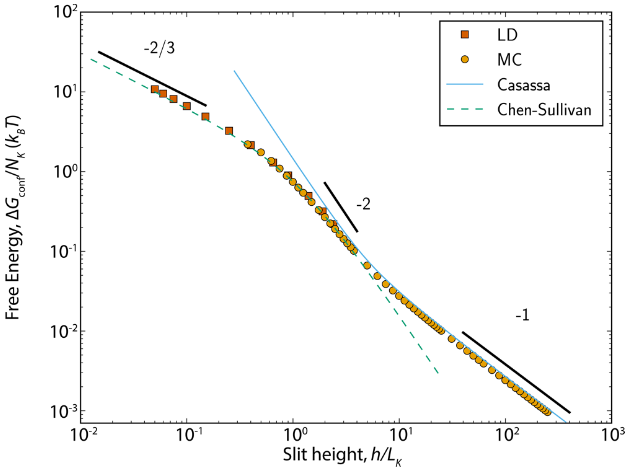


Figure 4. ΔG_{conf} per number of Kuhn lengths, N_K, versus slit height per Kuhn length, h/L_K, as obtained by the Monte Carlo (circles) and Langevin dynamics (squares) simulations. The three scalings of −2/3, −2, and −1 as well as all transitions between them are obtained for the pUC19 polymer model.

The Monte Carlo algorithm used in this work becomes very inefficient at tight confinement. Hence, we also performed Langevin dynamics (LD) simulations to explore the free energy of confinement at very small heights. In these simulation, we set the bead size, σ, to be 5 nm, which is on the order of the effective width of dsDNA. From this, the persistence length is set to 10σ, and the polymer is composed of 183 beads to give a contour length corresponding to pUC19. The polymer is initialized between two walls, and monomer–wall interactions are defined by standard coarse-grained potentials (see Supporting Information for details). The free energy of confinement is measured by recording the force exerted on the walls by the polymer at different slit heights. This data is then integrated to determine the free energy.

Figure 4 displays the LD and MC data across a wide range of slit heights. Note that the LD and MC data overlap for a significant range of heights and are found to be in excellent agreement with each other. Further, the LD simulations go to very tight confinement where pure Odijk scaling (exponent of −2/3) is obtained. Conversely, the MC calculations extend to very large heights, and the bulk confinement scaling of −1 is clearly recovered. The two numerical approaches overlap in the moderate confinement regime. Although the range is quite small for this particular polymer model, a steeper scaling of around −2 is apparent from both approaches. The Chen–Sullivan and Casassa transitions are also plotted in Figure 4, and the agreement between both calculations and these expressions is excellent. The combination of these two simulation approaches for the free energy of confinement for a semiflexible polymer thus starts in the pure scaling Odijk scaling of tight confinement, follows the Chen–Sullivan transition to moderate confinement, and finally transitions to bulk scaling in accordance with the Casassa formulation. To the best of our knowledge, this represents the first time that this complete mapping of $\Delta G(h)$ has been performed.

Returning to the experimental data shown in Figure 3, the points at the lowest h values lie in the CS transition where $h \lesssim L_K$ (and thus not $h \ll L_K$). For $h > L_K$, the data are transitioning from a scaling close to Odijk ($G_{conf} \sim h^{-2/3}$) toward moderate-confinement scaling ($G_{conf} \sim h^{-2}$). As pure moderate-confinement behavior is not expected until $h \gg L_K$, the CS transition is quite broad with a gradual approach to h^{-2} scaling. While we observe a shallower slope in the dependence of G_{conf} on h at $h < L_K$ than prediced by Chen and Sullivan, we cannot conclude that the Chen−Sullivan prediction is in error, as experimental uncertainty and artifacts may be significant enough to explain the discrepancy.

The difference between experimentally measured G_{conf} and simulated and theoretical G_{conf} at very low h ($<L_K$) may owe to a combination of (1) chamber characterization being less accurate near the point of contact between the two surfaces of the CLiC chamber (see Supporting Information), (2) statistical error in the small number of particles counted at very low heights, and (3) attraction of DNA to the glass surface, which would be the most significant at chamber locations with high surface-to-volume ratios. The wall attraction effects are likely contributors, as the attraction would result in the observed decrease in free energy. Additionally recent molecular simulations have shown weak DNA−silica attraction despite the electrostatic repulsion between negative silanolate moieties and the backbone phosphates and in the absence of chaotropic salts.[31] Moreover, the DNA in our experiments is likely to be attracted still more to the glass surface, as YOYO-1 is positively charged. These factors would also explain why the simulations agree with CS but not with experiments, as experimental error in height determination, greater statistical error at low h, and surface effects are not included in the simulation model. Surface interactions beyond excluded-volume repulsion have received little attention from theorists working under an Odijk theory paradigm, although our results suggest that they may be important in experiments taking place in the Odijk regime or the transition from moderate confinement thereto. Future research could incorporate results from size-exclusion chromatography theory that does consider these interactions (e.g., refs 32 and 33) and might bring simulations and a modified Chen−Sullivan theory into closer agreement with our experimental results.

Because for pUC19 $2R_g$ is little larger than L_K, the range of h in which pure, nontransitional moderate-confinement behavior is observed is small. This moderate-confinement window is defined by the range of h in which both CS and Casassa theory show approximately h^{-2} scaling, roughly $2R_g < h < 4R_g$. Above this window, the free energy is expected to depart the CS curve and approach the $G_{conf} \sim h^{-1}$ scaling predicted by Casassa in the limit of very large h. Indeed, our data depart the CS curve at a height predicted by Casassa, roughly $4R_g$. To the best of our knowledge, this result represents the first experimental and simulation test of the predicted Casassa transition to bulk for a semiflexible polymer.

We note that the measured free energy slope appears slightly shallower than the theoretical slope at large h. However, one cannot infer from our data that the Casassa theory is incorrect. The deviation is likely due to biases associated with two kinds of systematic experimental errors present for large heights. First, the free energy of confinement in log-space is especially sensitive to the bulk pUC19 concentration, which depends on a one-parameter fit of the combined data sets (see Supporting Information). Indeed, the larger error bars in the free energy at large h reflect reduced confidence in the fitted value of G_{conf} at large h (Figure 3). Further, particles can be identified and tracked only if they stay reasonably within the depth of field of our optical system: at higher h, it is increasingly possible for particles to make excursions out of this range, allowing for small fractions of their trajectories to be lost and thus G_{conf} overestimated (Supporting Information).

■ CONCLUSION

In conclusion, we have directly measured the free energy of confinement for semiflexible polymers in a slit-like geometry across a wide range of slit heights by counting DNA molecules within a CLiC chamber. Experimental results and Monte Carlo simulations connect confinement regimes described by the theories of Odijk (strong), Chen and Sullivan (strong-to-moderate), and Casassa (moderate-to-bulk). This comprehensive validation indicates that eqs 3 and 4 provide accurate and versatile predictions even beyond the polymer length scales for which the Chen−Sullivan model was derived. The experimental and simulation results, methods, and instrumentation delivered by this paper provide crucial information for modeling and understanding the behavior of polymers in a range of nanofluidic devices and facilitate efficient loading of and control over biopolymer samples within next-generation nanotechnology devices.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.macromol.6b01805.

> Detailed description of flow cell and sample preparation; CLiC microscopy setup; data collection; data analysis; and simulation techniques used (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: hendrick.dehaan@uoit.ca (H.W.d.H.).
*E-mail: sabrina.leslie@mcgill.ca (S.R.L.).

ORCID

Albert Kamanzi: 0000-0002-6790-1140

Author Contributions

J.S.L. and A.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Berard, D. J.; Michaud, F.; Mahshid, S.; Ahamed, M. J.; McFaul, C. M. J.; Leith, J. S.; Bérubé, P.; Sladek, R.; Reisner, W.; Leslie, S. R. Convex lens-induced nanoscale templating. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13295−300.

(2) Doyle, P. S.; Bibette, J.; Bancaud, A.; Viovy, J.-L. Self-assembled magnetic matrices for DNA separation chips. *Science* **2002**, *295*, 2237.

(3) Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; Gullans, S. R.; Gilmanshin, R. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res.* **2004**, *14*, 1137−46.

(4) Das, S. K.; Austin, M. D.; Akana, M. C.; Deshpande, P.; Cao, H.; Xiao, M. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Res.* **2010**, *38*, e177.

(5) Dimalanta, E. T.; Lim, A.; Runnheim, R.; Lamers, C.; Churas, C.; Forrest, D. K.; de Pablo, J. J.; Graham, M. D.; Coppersmith, S. N.; Goldstein, S.; Schwartz, D. C. A microfluidic system for large DNA molecule arrays. *Anal. Chem.* **2004**, *76*, 5293−301.

(6) Cao, H.; Yu, Z.; Wang, J.; Tegenfeldt, J. O.; Austin, R. H.; Chen, E.; Wu, W.; Chou, S. Y. Fabrication of 10 nm enclosed nanofluidic channels. *Appl. Phys. Lett.* **2002**, *81*, 174−176.

(7) Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; Gullans, S. R.; Gilmanshin, R. DNA Mapping Using Microfluidic Stretching and Single-Molecule Detection of Fluorescent Site-Specific Tags. *Genome Res.* **2004**, *14*, 1137−1146.

(8) Jo, K.; Dhingra, D. M.; Odijk, T.; de Pablo, J. J.; Graham, M. D.; Runnheim, R.; Forrest, D.; Schwartz, D. C. A single-molecule barcoding system using nanoslits for DNA analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2673−2678.

(9) Stavis, S. M.; Geist, J.; Gaitan, M.; Locascio, L. E.; Strychalski, E. A. DNA molecules descending a nanofluidic staircase by entropophoresis. *Lab Chip* **2012**, *12*, 1174−1182.

(10) Odijk, T. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules* **1983**, *16*, 1340.

(11) Klotz, A. R.; Duong, L.; Mamaev, M.; de Haan, H. W.; Chen, J. Z. Y.; Reisner, W. W. Measuring the Confinement Free Energy and Effective Width of Single Polymer Chains via Single-Molecule Tetris. *Macromolecules* **2015**, *48*, 5028−5033.

(12) Chen, J. Z.; Sullivan, D. Free energy of a wormlike polymer chain confined in a slit: crossover between two scaling regimes. *Macromolecules* **2006**, *39*, 7769−7773.

(13) Casassa, E. F. Equilibrium distribution of flexible polymer chains between a macroscopic solution phase and small voids. *J. Polym. Sci., Part B: Polym. Lett.* **1967**, *5*, 773−778.

(14) Odijk, T. Theory of lyotropic polymer liquid crystals. *Macromolecules* **1986**, *19*, 2313.

(15) Burkhardt, T. Free energy of a semiflexible polymer in a tube and statistics of a randomly-accelerated particle. *J. Phys. A: Math. Gen.* **1997**, *30*, L167−L172.

(16) Casassa, E. F.; Tagami, Y. An equilibrium theory for exclusion chromatography of branched and linear polymer chains. *Macromolecules* **1969**, *2*, 14.

(17) De Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.

(18) Tree, D. R.; Muralidhar, A.; Doyle, P. S.; Dorfman, K. D. Is DNA a good model polymer? *Macromolecules* **2013**, *46*, 8369−8382.

(19) Wang, Y.; Peters, G. H.; Hansen, F. Y.; Hassager, O. Equilibrium partitioning of macromolecules in confining geometries: Improved universality with a new molecular size parameter. *J. Chem. Phys.* **2008**, *128*, 124904.

(20) Tree, D. R.; Reinhart, W. F.; Dorfman, K. D. The Odijk Regime in Slits. *Macromolecules* **2014**, *47*, 3672−3684.

(21) Reisner, W.; Morton, K. J.; Riehn, R.; Wang, Y. M.; Yu, Z.; Rosen, M.; Sturm, J. C.; Chou, S. Y.; Frey, E.; Austin, R. H. Statics and Dynamics of Single DNA Molecules Confined in Nanochannels. *Phys. Rev. Lett.* **2005**, *94*, 196101.

(22) Bonthuis, D. J.; Meyer, C.; Stein, D.; Dekker, C. Conformation and Dynamics of DNA Confined in Slitlike Nanofluidic Channels. *Phys. Rev. Lett.* **2008**, *101*, 108303.

(23) Tang, J.; Levy, S. L.; Trahan, D. W.; Jones, J. J.; Craighead, H. G.; Doyle, P. S. Revisiting the Conformation and Dynamics of DNA in Slitlike Confinement. *Macromolecules* **2010**, *43*, 7368−7377.

(24) Cifra, P. Weak-to-strong confinement transition of semi-flexible macromolecules in slit and in channel. *J. Chem. Phys.* **2012**, *136*, 024902.

(25) Dai, L.; Jones, J. J.; van der Maarel, J. R. C.; Doyle, P. S. A systematic study of DNA conformation in slitlike confinement. *Soft Matter* **2012**, *8*, 2972−2982.

(26) Leslie, S. R.; Fields, A. P.; Cohen, A. E. Convex lens-induced confinement for imaging single molecules. *Anal. Chem.* **2010**, *82*, 6224−9.

(27) Berard, D.; McFaul, C. M. J.; Leith, J. S.; Arsenault, A. K. J.; Michaud, F.; Leslie, S. R. Precision platform for convex lens-induced confinement microscopy. *Rev. Sci. Instrum.* **2013**, *84*, 103704.

(28) Teraoka, I. *Polymer Solutions: an Introduction to Physical Properties*; John Wiley: New York, 2002.

(29) Ahamed, M. J.; Mahshid, S.; Berard, D. J.; Michaud, F.; Sladek, R.; Reisner, W. W.; Leslie, S. R. Continuous Confinement Fluidics: Getting Lots of Molecules into Small Spaces with High Fidelity. *Macromolecules* **2016**, *49*, 2853−2859.

(30) Sbalzarini, I.; Koumoutsakos, P. *J. Struct. Biol.* **2005**, *151*, 182−195.

(31) Shi, B.; Shin, Y. K.; Hassanali, A. A.; Singer, S. J. DNA Binding to the Silica Surface. *J. Phys. Chem. B* **2015**, *119*, 11030−11040.

(32) Cifra, P.; Bleha, T. Simulations of Partitioning in Size Exclusion Chromatography. *Int. J. Polym. Anal. Charact.* **2001**, *6*, 509−520.

(33) Cifra, P.; Bleha, T. Partition Coefficients and the Free Energy of Confinement from Simulations of Nonideal Polymer Systems. *Macromolecules* **2001**, *34*, 605−613.

DOI: 10.1021/acs.macromol.6b01805
Macromolecules 2016, 49, 9266−9271

E

APPENDIX: Spin-induced Macromolecular Spooling

Under revision (Physical Review X)

Tyler N. Shendruk, **David Sean**, Daniel Berard, Julian Wolf, Justin Dragoman,
Sophie Battat, Gary W. Slater and Sabrina R. Leslie

The following manuscript was submitted to Physical Review X. Referee comments were mostly positive and a second revision is in production.

Spin-induced macromolecular spooling

Tyler N. Shendruk*,[1,†] David Sean*,[2] Daniel Berard*,[3] Julian Wolf,[3] Justin Dragoman,[3] Sophie Battat,[3] Gary W. Slater,[2] and Sabrina R. Leslie[3,‡]
[1] *The Rudolf Peierls Centre for Theoretical Physics,*
University of Oxford, 1 Keble Road, Oxford, OX1 3NP, United Kingdom
[2] *Department of Physics, University of Ottawa, 150 Louis-Pasteur, Ottawa, K1N 6N5, Canada*
[3] *Department of Physics, McGill University, Montreal, H3A 2T8, Canada*

(Dated: August 1, 2016)

Genetic information is stored in a linear sequence of base-pairs; however, thermal fluctuations and complex DNA conformations such as folds and loops make it challenging to order genomic material for *in vitro* analysis. In this work, we discover that spin-induced spooling of DNA around a rotating microwire can monotonically order genomic bases, overcoming this challenge. We use single-molecule fluorescence microscopy to directly visualize long DNA strands deforming and elongating in shear flow near a rotating microwire, in agreement with numerical simulations. For the case of untethered DNA, when the rotation rate exceeds a critical value, we observe the Weissenberg effect at the single-molecule level for the first time. While untethered DNA is observed to elongate substantially, in agreement with our theory and numerical simulations, full extension of DNA becomes possible by introducing tethering. For the case of tethered DNA, we show that slowly increasing the rotation rate can deterministically spool a substantial portion of the DNA chain into a fully stretched, single-file, and overlap-free conformation. The fraction of genetic information sequentially ordered on the microwire surface increases with the DNA contour length, despite the increased entropy. This ability to handle genomic-length DNA is in contrast to modern DNA sample preparation technologies for sequencing and mapping, which are typically restricted to comparatively short DNA strands resulting in challenges in reconstructing the genome. Thus, in addition to discovering new spin-induced macromolecular dynamics, this work inspires new approaches to handling genomic-length DNA strands.

I. INTRODUCTION

Although the double-helix structure of DNA stores genetic information linearly as a sequence of bases at the molecular level, entropy randomizes the three-dimensional conformations of DNA polymers in free solution, such that accessing long genomic sequences is exceedingly challenging. Eukaryotic cells actively overcome genomic entropy by wrapping DNA around histones, which are then further dynamically twisted and coiled to form highly condensed and ordered chromosomes [1]. By wrapping DNA in this manner, genomic data is packaged, condensed, and regulated in mitotic chromosomes, in a manner that is unrivaled by physical technologies.

To manipulate and study DNA, physical approaches have been developed to spatially arrange DNA in a controlled format, including magnetic and optical traps [2, 3], microfluidic approaches [4–7], and nanoconfinement approaches [8], such as Convex Lens-induced Confinement (CLiC) [9] or nanopore confinement [10]. However, *in vivo* manipulation of genomic-length biopolymers remains an outstanding challenge that, with existing approaches, becomes more difficult with increasing strand length.

In this work, we untangle and order individual DNA molecules in long robust sections, without breakage. Using fluorescence imaging, we demonstrate that DNA strands deform and elongate in shear flow near a rotating microwire when the rotation rate exceeds a critical value. Our experiments produce the first single-molecule realization of the Weissenberg effect.

While single strands of untethered DNA are observed to elongate substantially in agreement with scaling theory and simulations, full DNA extension becomes obtainable by combining tethering with this DNA spooling approach. We find that tethering one end of the DNA to the rotating microwire enhances extension, and produces novel "shofar" conformations. By slowly increasing the experimentally realizable rotation rate above an additional critical value, a substantial portion of the tethered DNA is spooled into a fully stretched, single-file, and overlap-free curvilinear conformation. Since the conformation consists of a "shofar"-type tail and a base-ordered stem, we refer to this as a "French-horn" conformation. The fraction of bases in the fully ordered stem of the French-horn conformation is found to increase with strand length.

This organizational capability holds promise for enabling new insights into extended "repeat-regions" in genomes, which are challenging to prepare and sequence in single continuous units. Current sequencing technologies typically rely on fracturing the genome into short fragments prior to readout; however, repeat-regions are challenging to identify and reassemble, obscuring potentially important genetic information [11]. Using our DNA

---

*T.N.S., D.S. and D.B. contributed equally to this work.
†Corresponding author: `tyler.shendruk@physics.ox.ac.uk`
‡Corresponding author: `sabrina.leslie@mcgill.ca`

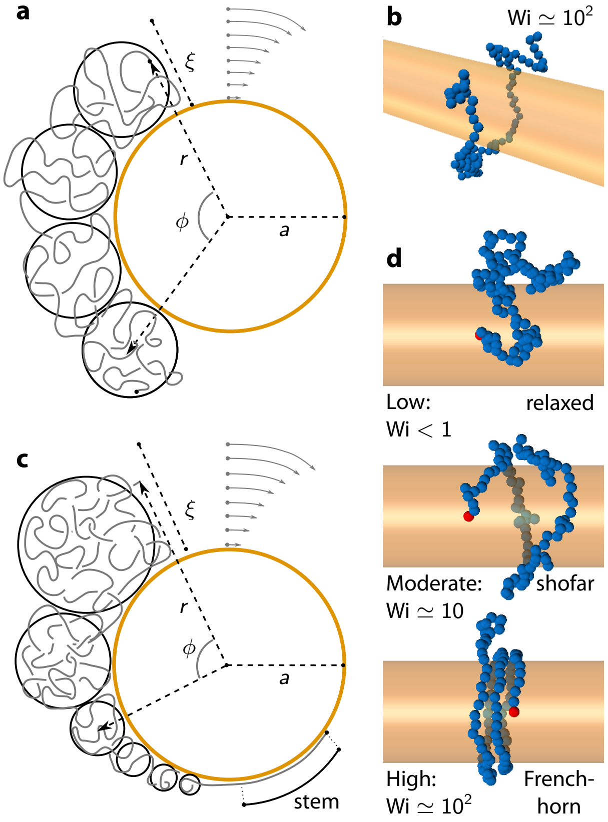


FIG. 1: **a.** The untethered DNA system: The rotating cylinder of radius a generates a flow profile $v_\phi(r)$, which shears the DNA and deforms it into a procession of identical blobs of diameter ξ. **b.** Simulation snapshot portraying a typical steady-state conformation of an untethered DNA chain. **c.** The tethered system with the French-horn DNA conformation: The blob size ξ is a function of angle ϕ from the untethered end. For sufficiently large Weissenberg numbers Wi, a portion of the DNA is fully stretched into a stem section, while the rest is in a horn conformation. For smaller Wi there is no stem and the polymer is in the shofar conformation. **d.** Simulation snapshots portraying typical steady-state tethered conformations. Increasing Wi changes the state from relaxed, to shofar, to French-horn conformations.

spooling approach, repeat-regions could be sequentially ordered and maintained on the surface of the rotating microwire to enable downstream readout over large genomic distances.

A. Spin-induced Spooling

The physical mechanism for compactly wrapping stretched DNA around a spinning microwire is presented in Fig. 1a for untethered DNA. The rotation drives an azimuthal flow, which shears the deformable DNA and draws it towards the surface of the microwire. When the DNA is untethered, the biopolymer elongates into a series of blobs and the system essentially generates a single-molecule equivalent of the Weissenberg effect. While the rotation rates required to fully stretch untethered DNA are very high, when one end is tethered to the rotating microwire, moderate rotation rates are sufficient to fully stretch significant portions of the DNA (Figs. 1c and 1d). The stretched DNA is tautly wrapped and unraveled into a coordinated single-file conformation. By gradually increasing the rotation rate, the blob size can become comparable to the DNA Kuhn length, while chain-breaking events can be avoided and the excluded volume interactions between the blobs ensure that overlaps are rare. In this way, single molecules of DNA can be organized into single-file conformations, deterministically ordering the genetic information on a cylindrical surface.

In prior work using planar geometries, DNA and other long macromolecules have been demonstrated to deform when subjected to experimentally achievable flows [12, 13]. For example, a wall-tethered polymer subject to sufficient shear rates deforms into a string of blobs [14, 15]. The physical mechanisms leveraged in this study can be understood by considering the deformation of untethered DNA. We thus consider untethered strands, after first introducing our methods, before subsequently analyzing tethered chains.

II. METHODS

When a cylindrical microwire of radius a is rotated with rate Ω, the no-slip boundary condition generates a flow profile $v_\phi(r) = \Omega a^2/r$ and a rapidly decaying shear rate $\dot{\gamma}(r) = -\Omega(a/r)^2$ at a distance r from the centre of the cylinder (Fig. 1). Each strand of DNA is composed of N Kuhn segments of length $b \approx 100\,\text{nm}$ and characteristic relaxation time $\tau_b = \eta b^3/k_\text{B}T \approx 2.4 \times 10^{-4}\,\text{s}$ in a solvent with dynamic viscosity η and thermal energy $k_\text{B}T$. For a chain in a good solvent with a Flory exponent $\nu \approx 3/5$, unperturbed DNA has a relaxation time $\tau \simeq \tau_b N^{3\nu}$ and a corresponding undeformed radius of gyration $R_{g0} \simeq bN^\nu$.

By suspending a microwire directly above a coverslip (Fig. 2a), submerging it in a drop of solution containing DNA, and gradually increasing its rotation rate, we observe the dynamics of DNA in a rotationally-induced shear flow. Far from the slowly rotating wire, the shear is small enough that untethered DNA appears relaxed and diffusive. Nearer to the microwire, the flow dominates and the DNA is advected. Some polymers, which are initially distributed evenly throughout the solution, are drawn towards the rotating microwire. This radial migration across streamlines towards the microwire arises from the combined effects of the hydrodynamic interactions with the wire, the nonhomogeneous flow, and the decrease of diffusivity with stretching [16]. Within the advection regime, many strands are observed to be elongated by the shear flow (Fig. 2b). The Weissenberg number $\mathrm{Wi} = \dot{\gamma}\tau$ expresses the competition between the characteristic shear forces and the polymer's intrinsic relaxation time τ. The shear deforms the DNA when $\mathrm{Wi} > 1$ (Fig. 2b and c). By substituting in the relaxation time and shear rate, we find that deformation occurs for $r/a \lesssim (\Omega\tau_b)^{1/2}\, N^{3\nu/2}$. This hints at the benefit of our technique when dealing with genome-length DNA, since the minimum rotation rate required to achieve deformation is $\Omega^*_{\mathrm{Wi}}\tau_b \simeq N^{-3\nu}$, which drops rapidly with DNA length.

A. Apparatus

The apparatus used to perform the experiments is shown in Fig. 3. A 50 µm diameter tungsten wire (Malin co.) is held and spun at both ends by two stepper motors (Pololu Robotics & Electronics part 1204). Each end of the wire is held within a short piece of fluorinated ethylene propylene (FEP) microfludic tubing with an inner diameter of 50 µm (IDEX Health & Science), each of which is mounted to one of the stepper motors using couplers machined from aluminum. The small inner diameter of the tubing serves to reduce microwire wobble. The mounting points of the wires do not exactly align, which results in minor wobbling about the rotation axis.

Each stepper motor is mounted on a z-axis-stage (Thorlabs part MS1S) which allows its height to be varied. This allows the vertical position of the wire above the sample coverslip, as well as the angle of the wire to the horizontal, to be set precisely. One of the stepper motors is also mounted on a horizontal stage which allows the distance between the motors to be varied, so that a controlled tension can be applied to the microwire. The motors and stages are mounted to a base machined from aluminum. The rotation of the motors is controlled by an Arduino microcontroller board.

The apparatus mounts on an inverted fluorescence microscope (Nikon Ti-E), and experiments are performed with a 60× NA 1.0 water immersion objective (Nikon CFI Plan Apo VC 60XWI) and an Andor iXon3 EMCCD camera at an EM gain of 270. In order to observe the DNA, the YOYO-1 dye (Section II B) is excited using 1.5 mW from a 488 nm Coherent Sapphire laser.

B. DNA Preparation

For untethered experiments, T4-phage DNA (166 kbp, New England Biosciences) is used at a concentration of 0.1 µg/mL. The DNA are stained with YOYO-1 fluorescent dye (Life Technologies) at a ratio of 1 dye molecule per 10 base-pairs. YOYO-1 increases the DNA contour length at this staining ratio from 56.4 µm to 65(2) µm for T4-DNA [17]. Experiments are carried out in 0.5× tris-borate-EDTA (TBE) buffer, with 3 % β-mercaptoethanol (BME) added as an anti-photobleaching agent. Before beginning untethered experiments, the microwire is first passivated using a solution of 10 % 55 kDa polyvinylpyrrolidone (PVP) in 0.5X

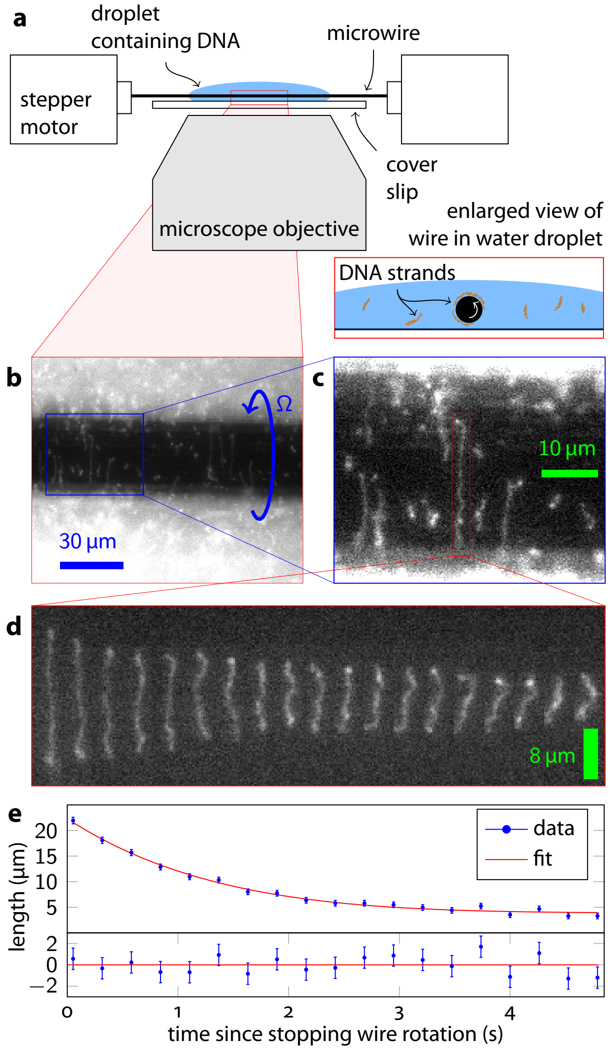


FIG. 2: The apparatus for untethered strands. **a.** Schematic representation of the apparatus, described in detail in the text. **b.** Sample image of a 50 µm-diameter tungsten wire (dark horizontal line) rotating in a dilute solution of T4-phage DNA (166 kbp). The contour length of YOYO-1 stained T4-DNA is (65 ± 2) µm. Strands of DNA are drawn towards the wire and elongated by the shear flow. **c.** Image of the microwire 1 s after it has stopped rotating: DNA strands have begun to relax to more entropic states. **d.** Conformational relaxation of a single DNA strand in the period of time immediately after the microwire has stopped rotating (see Movie S1). **e.** Quantitative measurements of the extension of a single DNA strand during relaxation, along with the fit and associated studentized residuals, used to determine the fully extended length of the polymer ($t = 0$).

TBE to reduce sticking of DNA to the glass coverslip and tungsten microwire.

For a solvent viscosity similar to that of water, the bulk relaxation time and radius of gyration of T4-DNA are $\tau = 9.69\,\mathrm{s}$ and $R_{g0} = 1.46\,\mu\mathrm{m}$ [18]. The goodness of solvent for T4-DNA in 0.5× TBE is $\nu = 0.558$ [19].

C. Measuring Extension

To accurately measure the end-to-end lengths L of the untethered DNA in the apparatus, the relaxation from the extended state is fit. The microwire rotates, extending the near-wire polymers, and then the rotation is suddenly halted. The DNA strands relax to a more highly entropic state with decreased extension and the fully extended length $L(t=0)$ is extracted is obtained from [20]

$$L^2\,(t) - 4R_{g0}^2 = A\,Nb\,e^{-t/\tau}, \tag{1}$$

where t is the time since the rotation of the microwire has stopped and A is a proportionality constant that is allowed to vary when fitting. This function is fit to the relaxation of each polymer, and the fully extended length $L(t=0) = \sqrt{4R_{g0}^2 + A\,Nb}$ is extracted from the fit (Fig. 2e).

Each individual measurement of the elongation length of the DNA strand at time t requires locating the positions of its two ends. The uncertainty in the measurement of each of these positions is dominated by the pixel size of the camera, corresponding to 267 nm. The uncertainty in each measurement is then $\delta L = 533\,\mathrm{nm}$, while the uncertainty in the fully extended length of the polymer is extracted from the fit to its relaxation. There is large variance in the extent of elongation as expected from results on tumbling DNA in simple shears [21, 22].

D. DPD Algorithm

The spin-induced macromolecular spooling does not rely on chemical details of DNA—it is a non-specific physical phenomenon. For this reason, a coarse-grain Dissipative Particle Dynamics (DPD) numerical model is used to simulate generic DNA molecules to extend our understandin of DNA dynamics in three dimensional azimuthal flows at various rotation rates with hydrodynamic interactions.

Both DNA monomers and fluid particles are represented by soft particles of mass m and size σ. The i^{th} particle is described by its coordinates in continuous-space, which evolve in response to inter-particle forces with other particles (labeled j). In DPD, these are commonly pair-wise conservative forces $\vec{F}_{ij}^{C}$ with a soft repulsion with a cut off r_c and strength a_{ij}, dissipative forces $\vec{F}_{ij}^{D}$ with inter-particle frictional drag γ_{DPD}, and random forces $\vec{F}_{ij}^{R}$ due to the thermal energy $k_{\mathrm{B}}T$. An extended discussion of these forces can be found in Ref. [23].

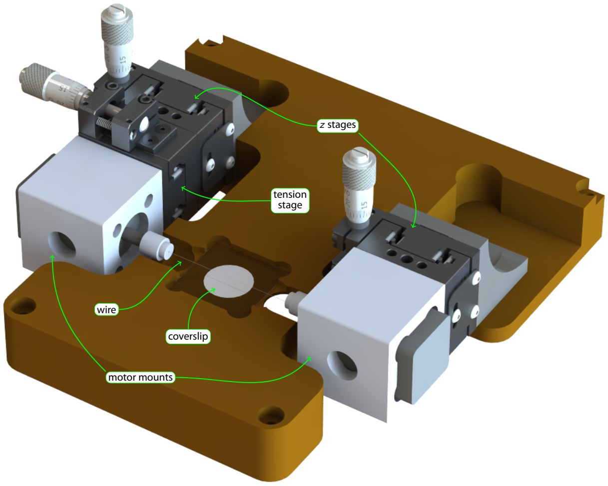


FIG. 3: The apparatus used to collect data on the spin-induced spooling of DNA. The microwire is situated directly above the cover slip using the z stages of the two motor mounts and the tension stage is adjusted so that the microwire is taut. Solution containing DNA is dropped onto the cover slip such that a portion of the wire is submerged. The motors are driven simultaneously by an Arduino microcontroller.

The number density of DPD beads in the simulations is ρ_f and through the fluctuation-dissipation theorem the drag is $\gamma_{\mathrm{DPD}} = \sigma^2/\left(2k_{\mathrm{B}}T\right)$. The simulation units are $[m=1; r_c=1; k_{\mathrm{B}}T=1]$ and we choose $\sigma = 3$, $\gamma_{\mathrm{DPD}} = 4.5$, $\Delta t = 0.01$, $\rho_f = 3$, and $a_{ij} = 25$.

The DPD beads that form the DNA strand have an additional conservative bonding force to sequentially polymerize $N = 100$ monomers. For this, we opt for the widely used FENE spring potential U_{FENE}, in conjuction with the purely repulsive truncated Lennard-Jones potential U_{WCA} [23]. We note that the both U_{FENE} and U_{WCA} are only applied between consecutive coarse-grained beads.

The DPD repulsion suffices to model long-ranged excluded-volume interactions, which yield the expected size exponent for a real polymer in a good solvent[24]. This allows us to obtain a monomer relaxation time $1 < \tau_b < 0.1$ [24], which we approximate as $\tau_b \approx 0.3$. The presence of the DPD fluid ensures that hydrodynamic interactions between monomers are reproduced, leading to Zimm dynamics. The bulk relaxation time of DPD chains kept in the vicinity of the surface of the microwire by tethereing in quiescent solvent is measured to be $\tau = 920$.

We model the microwire as a semi-infinite cylindrical surface of radius $a = 5$ composed of three layers of DPD beads, which rotate with a fixed angular velocity. A concentric cylinder of radius $a_{\mathrm{out}} = 35$ provides a finite size to the system and is constituted by a single layer of DPD beads that rotates with the theoretically expected speed $v_\phi\left(r = a_{\mathrm{out}}\right) = \Omega a^2/a_{\mathrm{out}}$. The resulting shears match the theoretical expectations except very near the microwire surface, where minor packing and slip-effects cause the flow profile to be over predicted. We apply periodic boundary conditions along the microwire's axial direction with a system thickness of 15. We observe that the soft nature of the DPD beads can lead to slight polymer penetration into the inner rod when the rotation rate is sufficiently high. To reduce this effect, we add a smooth mathematical cylindrical surface that is transparent to the DPD beads (hence does not explicitly affect the flow) but unto which we apply U_{WCA} for every polymer bead. With the surface characterized by a radius $a - 1/2$, the polymer feels a steric repulsion and does not penetrate the surface of the nominal rod.

The DNA model is initialized tightly wrapped around the cylinder and left to equilibrate. By noting the monomer position in a cylindrical coordinate system $\vec{r}_i(r_i, \phi_i, z_i)$, it is possible to find the instantaneous span $L_s \equiv \langle r \rangle \Delta\phi = \langle r \rangle (\phi_{\max} - \phi_{\min})$. A tightly wrapped polymer is thus initialized at a high value of L_s, which exponentially decreases to a steady state value (characterized by $\tau_{\exp}$). An exponential fit to Eq. 1 is conducted to determine the steady-state curvilinear extension L. The blob size is measured directly as the average monomer distance from the cylinder surface $\xi = \langle r \rangle - a$. As in experiments, only DNA strands in the elongated phase of the tumbling cycle [22] are analyzed.

Although we find that the monomers are driven towards the cylinder surface as expected, when the rotation rate is low, untethered polymers can sometimes escape into the bulk. Although re-capture events are observed, escaped polymers are not analyzed. As in simple shear, tumbling dynamics [25] and U-turn conformations [26] are observed in spin-induced macromolecular spooling. This sets a lower bound to the rotation rate since the DNA commonly escapes from the surface of the rotating wire at slow rotation rates.

III. UNTETHERED DNA

A. Weakly Extended Untethered DNA

The extended lengths L of elongated polymers are measured at a number of rotation rates (Fig. 4). Below $\Omega \approx 7\,\text{rpm}$, some strands are drawn towards the rotating microwire; however, their elongation is weak. In this weak elongation regime ($\text{Wi} \gtrsim 1$), the DNA behaves as a thermal spring subjected to hydrodynamic stretching forces. The incident shear rate $\dot{\gamma} = \Omega a^2 / (a + \xi)^2 \simeq \Omega$ enacts a characteristic dimensionless drag $Fb/k_{\text{B}}T \simeq \Omega \tau_b N^{2\nu}$ on the weakly deformed DNA. Since the dimensionless thermal spring force scales as $\simeq L/(Nb)$, we predict that in the weak elongation regime

$$L/(Nb) \simeq \Omega \tau_b N^{2\nu}, \tag{2}$$

which can also be written $L/R_{g0} \simeq \text{Wi} N^{1-2\nu}$ and is independent of wire diameter.

The DNA extension increases rapidly with rotation rate as $L \sim \text{Wi} \sim \Omega$ in this regime and this is consistent with the low rotation rate data in Fig. 4. However, a crossover from an initially rapid rise at moderately large Wi to a much slower rise when $\text{Wi} \gg 1$ is expected; the experimental data shows a marked crossover from the weak-extension regime to a much lower scaling at high rotation rates.

B. Highly Extended Untethered DNA

When the rotation rate is increased further, the deformed DNA elongates into a string of blobs and aligns along the azimuthal streamlines at the surface of the rotating cylinder (Fig. 1b). Visual inspection of the elongated polymers reveals a uniform blob size ξ along each backbone in the highly extended regime (Fig. 2c-d). Tension is equally distributed throughout the chain and each blob has a relaxation time $\tau_\xi = \tau_b (\xi/b)^3$. Since shear and relaxation balance when the Weissenberg number for each blob is unity, the untethered blob size is

$$\xi/b \simeq (\Omega \tau_b)^{-1/3} \qquad \text{if } \xi \lesssim a. \tag{3}$$

This result can be written as $\xi/R_{g0} \simeq \text{Wi}^{-1/3}$.

The blob size is experimentally inaccessible due to the diffraction limited resolution. Therefore, we consult numerical experiments to verify this prediction. DPD simulations explicitly estimate the blob size ξ as the average distance of the monomers from the surface of the wire (Fig. 4 inset). For sufficient rotation rates, ξ scales with rotation rate in the manner predicted by Eq. 3. At low rotation rates, the DNA often diffuses into bulk solution from the wrapped state. This places a lower bound on the rotation rates that can be numerically investigated.

The curvilinear end-to-end distance of the highly deformed DNA on the wire surface is predicted to be

$$L/(Nb) \simeq (\Omega \tau_b)^{(1-\nu)/(3\nu)} \tag{4}$$

by considering the number of Kuhn segments in a single blob and the total number of blobs. This result is equivalent to $L/R_{g0} \simeq \text{Wi}^{(1-\nu)/(3\nu)}$. The extension obviously increases linearly with contour length Nb such that that longer chains stretch more easily than shorter chains, but the extension is only weakly dependent on rotation rate in this highly extended regime. Inserting $\nu \approx 3/5$ predicts a weak dependence on rotation rate ($L/R_{g0} \sim \text{Wi}^{2/9}$) for this regime, which stands in contrast to the initial linear dependence in Eq. 2 for weakly stretched chains.

At the highest experimentally achievable rotation rates in Fig. 4, the extension of the untethered DNA scales in accordance with Eq. 4. Due to limitations on the current apparatus, the microwire cannot be steadily driven above $\Omega \approx 130\,\text{rpm}$ without inducing significant wire wobble. The DPD data extend to higher Wi, and at moderate Wi

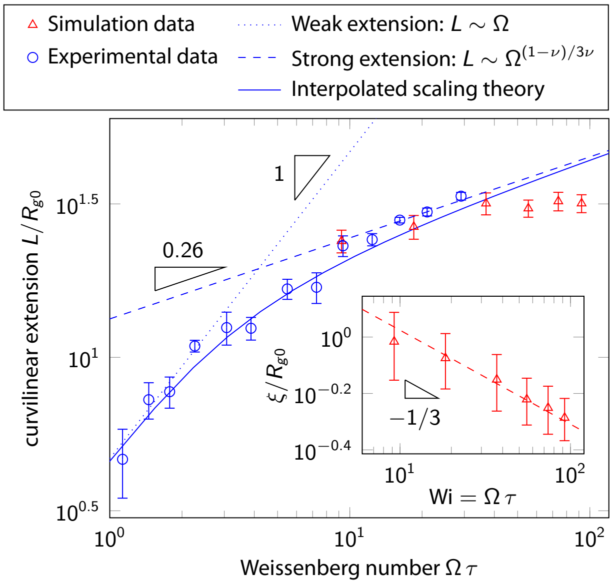


FIG. 4: Curvilinear end-to-end extension of untethered T4-DNA (166 kbp) and DPD polymers for varying rotation rates. The model-dependent shift between the high Weissenberg-number simulations and experiments is removed by setting the $\mathrm{Wi} = 10$ value to the fitted strong extension value. Dotted and dashed lines show the predicted scaling for weak extensions at low Wi (Eq. 2; $L \sim \mathrm{Wi}$) and large extensions at high Wi (Eq. 4; $L \sim \mathrm{Wi}^{(1-\nu)/3\nu} \sim \mathrm{Wi}^{0.26}$), respectively. **Inset.** Blob size ξ of the DPD polymer. Dashed line shows the scaling predicted by Eq. 3.

agree with the exponent predicted by Eq. 4 and observed experimentally before exhibiting a plateau. The plateau at higher Wi is likely a result of extension beyond the blob regime due to the finite length of the DPD chain compared to T4-DNA.

Although both the weak and the high extension regimes are evident in Fig. 4, relatively large rotation rates are required to reach the highly extended regime. While DNA with longer contour lengths will have larger Weissenberg numbers and so be further deformed at experimentally realizable rotation rates, this highlights the shortcoming of applying the method to untethered DNA. Because of the weak scaling in this highly extended regime, untethered DNA can easily be deformed but not fully extended. To obtain a fully-stretched chain would require $\xi \to b$ and, therefore, $\Omega^{*}_{\mathrm{full}} \simeq \tau_b^{-1} \approx 4000\,\mathrm{s}^{-1}$, which would be experimentally challenging. On the other hand, we next demonstrate that experimentally achievable rotation rates can fully stretch large portions of tethered DNA.

IV. TETHERED DNA

Untethered strands of DNA readily deform in the vicinity of the rotating microwire, but reaching the fully stretched limit is infeasible and fluctuations due to tumbling dynamics are significant. However, by anchoring one end of the DNA to the rotating microwire, a substantial portion of sufficiently long chains can be fully stretched.

For the tethered experiments, λ-phage DNA (48.5 kbp, New England Biosciences) is used, also at a concentration of 0.1 µg/mL and stained with YOYO-1 fluorescent dye as in the untethered case. Staining increases the λ-DNA contour length from 16.5 µm to 19.0(7) µm [17]. Custom Cy5-labeled oligos complementary to the 12-bp single-stranded overhangs at the ends of λ-phage DNA are added (IDT). The hydrophobic character of the Cy5 end-labels results in a significant increase in surface-ahesion. The glass coverslip is passivated with PVP as in the untethered case, but the microwire is not. Tethered DNA molecules, for which one end has attached to the wire, are used for quantitative analysis of the tethered case. Since the hydrophobic Cy5-labels are positioned at the DNA ends, immobilization occurs at the ends.

Similar to the classical differences between a chain dragged by one end through a quiescent fluid and a wall-tethered chain subjected to shear [13, 27], the tethered system setup differs from a chain attached to a filament [28] in that it is the shearing flow due to rotation that leads to deformation. At low rotation rates ($\Omega < \Omega^{*}_{\mathrm{Wi}}$), the tethered DNA remains relaxed and simply rotates with the anchoring point. Simulation snapshots at low rotation rates (Fig. 1d; $\mathrm{Wi} \simeq 10^{-0.3}$) show that the tethered DNA molecules follow the rotation in a quasi-static manner. They remain in a thermalized state throughout the complete simulation runs because the shear is too low to stretch the DNA into an out-of-equilibrium conformation. The critical rotation rate for deformation is $\Omega^{*}_{\mathrm{Wi}}\tau_b \simeq N^{-3\nu}$, which is the same as in the untethered state.

A. Shofar Conformation

For greater rotation rates, the tethered polymer deforms into a series of blobs $\xi_i(\phi)$ that decrease in size as a function of the angle ϕ from the free end (Fig. 1c). Using DNA which is end-labeled with “sticky” Cy5 tags (*Materials and Methods*), tethering of polymers to the rotating microwire can be realized. At a rotation rate $\Omega \approx 130$ rpm a configuration of varying blob sizes is readily achievable and an example of the growth of the blob size away from the tethering point is clear in Fig. 5a.

The i^{th} blob is described by the radial distance $r_i = a + \xi_i$ and arc length $s_i = r_i\phi$. The drag force on each blob is $f_i \simeq \eta v_\phi(r_i)\,\xi_i$ but the tension on each segment is due to the total drag of all the preceeding blobs. The tension within the chain is thus $\sum_j f_j = \sum_j \eta v_\phi \xi_j \to$

$\int \eta v_\phi \xi \left(ds/\xi\right)$ and, equivalently through the definition of the blob size, it is also $k_\mathrm{B}T/\xi$. Therefore, the blob size as a function of angle ϕ from the free end is

$$\xi\left(\phi\right)/b \simeq \left(\frac{a}{b}\Omega\tau_b\phi\right)^{-1/2}. \tag{5}$$

This result reproduces the classic planar-shear result when the shear rate $\dot{\gamma}$ replaces Ω and the linear distance replaces the arc length $s = a\phi$ [27, 29]. If the planar case may be described as a "trumpet" conformation [30] then our experiments produce a cylindrical "shofar" conformation. Inspection of Fig. 1d shows that a moderate rotation rate is sufficient to deform DNA and wrap it around the microwire. There remain visible thermal fluctuations along the backbone, and the size of the blobs can be seen to increase with the distance to the tethering point, thus forming the shofar conformation.

Assuming the tension varies slowly along the chain, we can write the rate of change of the local extension s with segment index n counted from the free end to be $ds/dn \approx b\left(b/\xi\right)^{(1-\nu)/\nu}$ [27], or $ds/dn \approx \left(a+\xi\right)\left(d\phi/dn\right)$ by inverting Eq. 5. Assuming that $a \gg b$ and considering n sufficiently far from the free end, we equate the two forms and integrate to find

$$\phi\left(n\right) \sim \left(b/a\right)\left(\Omega\tau_b\right)^{(1-\nu)/(3\nu-1)} n^{2\nu/(3\nu-1)}. \tag{6}$$

Within the shofar conformation of blobs and sufficiently far from the free end, simulations show that the angle from the free end grows with segment index as described by Eq. 6 (Fig. 5b inset; dashed line). As expected, the scaling does not hold near the free end.

Further details about the conformation can be obtained from the blob size as a function of segment index from the free end. Substituted into Eq. 5, the angle $\xi\left(n\right)$ produces

$$\xi\left(n\right)/b \simeq \left(\Omega\tau_b n\right)^{-\nu/(3\nu-1)}, \tag{7}$$

As in the planar case, Eq. 5 predicts that the blob size diverges as $n \to 0$. Therefore, it is truncated at the first blob $\phi_1 \approx \xi_1/a$. This gives $\xi_1/b \simeq \left(\Omega\tau_b\right)^{-1/3}$, which is the same as the untethered blob size (Eq. 3). The free-end blob remains clearly identifiable at small n in measurements of the blob size (Fig. 5b). The blob size crosses to the shofar conformation that was observed in the inset and is seen to be well described by Eq. 7 (Fig. 5b; dotted line) for DNA segments in the mid-region of the chain. The blob size at the tethering point is predicted to be $\xi\left(N\right)/b \simeq \left(\Omega\tau_b N\right)^{-\nu/(3\nu-1)}$. However, as the tethered end is approached, the blob size as a function of index begins to saturate to the segment size for many of rotation rates in Fig. 5b. These curves represent a new conformation.

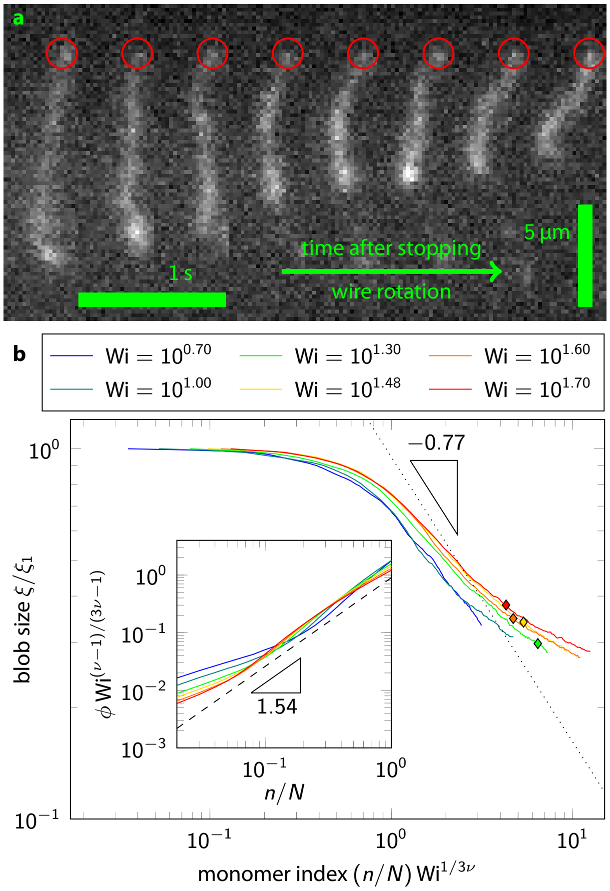


FIG. 5: Spooling tethered DNA strands. **a.** Conformation of a single λ-phage DNA strand in the period immediately after the microwire has stopped rotating at $\Omega = 130\,\mathrm{rpm}$, showing the relaxation of the polymer (see Supplemental Video 2). The end of the polymer that appears towards the top of the image is tethered to the wire. The blob size of the polymer is observed to increase steadily away from the tethering point, as predicted for a tethered polymer in the shofar configuration. **b.** French-horn conformations seen via DPD blob size ξ as a function of segment index from the free-end. Various rotation rates are collapsed according to Eq. 7 with blob size normalized by the free-end blob size ξ_1. Dotted line shows scaling prediction $\xi \sim n^{-\nu/(3\nu-1)} \sim n^{-0.77}$. Diamonds ($\diamond$) denote the predicted transition from shofar to stem at $n^* \simeq \left(\tau_b\Omega\right)^{-1}$. **Inset.** Wrapping angle from the free end, collapsed via Eq. 6. Dashed line shows the scaling $\phi \sim n^{2\nu/(3\nu-1)} \sim n^{1.54}$.

B. French-horn Conformation

If the rotation rate is increased to $\Omega^*_\mathrm{horn}\tau_b \simeq N^{-1}$ then the blob size at the tethering point is predicted to be equal to the minimum size b. For higher rotation rates ($\Omega > \Omega^*_\mathrm{horn}$), the segments near the tethering point are fully stretched and the chain adopts a "French-horn" conformation with a fully stretched "stem" near the tethering point followed by a shofar-shaped horn of blobs near the free end (Fig. 1c). The onset of this French-horn conformation occurs at substantially slower rotation rates than fully stretched conformations. The segment index where the transition from shofar to stem occurs is $n^* \simeq \Omega^*_\mathrm{full}/\Omega \simeq \left(\tau_b\Omega\right)^{-1}$, and so the fraction of segments in the fully stretched stem is

$$\lambda \simeq \frac{N-n^*}{N} = 1 - \left(N\Omega\tau_b\right)^{-1}. \tag{8}$$

Due to limitations on the current apparatus ($\approx 130\,\mathrm{rpm}$ maximum rotation rate), the stem has not yet been experimentally observed. Though the relaxation time of a single Kuhn segment ($\tau_b \sim 10^{-4}\,\mathrm{s}$) is quite short, N is large for genome-length DNA such that experimentally obtainable rotation rates are predicted to fully stretch significant portions of the polymer.

A rotation rate of $\Omega = 0.1$ suffices to wrap the majority of the monomers in a single-file manner in simulations (Fig. 1d; Wi $\simeq 10^2$). Here, the stem-state dominates with a small proportion of monomers forming a blob bigger than the monomer size at the free DNA extremity. In simulations, the stem portion does not reach a constant blob size cutoff due to the softness of the DPD beads and slight bond stretching, which results in a reduced but non-zero decrease in effective blob size (Fig. 5b). However, the predicted crossover from the shofar to the stem of the French-horn conformation at n^* is seen to approximate this point (Fig. 5b; $\diamond$).

Our results reproduced the French-horn conformation for tethered chains, providing confidence that a significant portion of the DNA can be expected to be fully stretched. To ensure that the fully stretched stem follows a deterministic screw-like, overlap-free organization, the rotation rate should be slowly increased. In this way, excluded volume effects between blobs act to prevent overlap at all times such that, when the stem forms, it is stretched into a non-overlapping single-file conformation, an important requirement for sequential ordering. While genetic information in the shofar portion of the DNA remains disordered in space, it is arranged cylindrically in the stem, just as linearly advancing music notes are punched on the cylinder of a music box.

V. DISCUSSION

Leading DNA sequencing technologies typically rely on reconstruction of genetic information from the analysis of many relatively short DNA fragments. Organization of single genome-length strands sequentially in space must overcome the high entropy associated with complex, folded polymer conformations. Spin-induced macromolecular spooling presents a new single-molecule manipulation concept, allowing DNA to be ordered into a curvilinear progression of base-pairs, arranged sequentially on the surface of a rotating microwire.

We observe untethered strands to be highly deformed, and enhanced extension to be possible with tethering. It is compelling that DNA tethered to the rotating microwire takes various previously-unseen conformational states, including shofar and French-horn conformations, at experimentally moderate rotations rates. The portion of the French-horn conformation that is in the fully extended stem state increases with both rotation rate and DNA length. In fact, by anchoring one end of a DNA strand to the microwire after utilizing DNA ligase to conjoin additional “attachment handles” to the free ends, future applications could tautly wrap entire DNA lengths into a deterministic single-file and unknotted conformation.

Furthermore, a subsequent experimental step could be implemented to deposit DNA onto the microwire, following its ordering on the surface, for storage and readout. In this way, the complete genetic sequence could be organized into a linear progression of base-pairs that are not susceptible to thermal fluctuations as random coils for subsequent readout. The genomic sequence could be stored on the wire for long periods and at high density, in analogy to punching sequential music notes on the surface of a music box cylinder.

While we have emphasized development of sample-preparation technology, biochemistry procedures for downstream sequencing of spooled DNA have yet to be developed. Future devices could utilize our work on spin-induced macromolecular spooling to develop technologies that act as front-end sample-handling instrumentation for next-generation sequencing devices, preparing and storing genome-length DNA strands in regular conformations. This could potentially allow each base to be interrogated by downstream devices without fragmenting strands or losing long-range genetic information such as insertions, deletions and the rearrangements of genes that are often associated with disease.

VI. POPULAR SUMMARY

Even though our genetic information is encoded in a linear sequence of bases in thread-like DNA, that thread can be tangled, just like a strand of yarn stuffed in your pocket. In order for scientists to read the genetic code, they must first straighten the DNA strands. Previous attempts to organize DNA amount to threading it through the eye of a needle, but this is not how you would organize yarn. Instead you would wrap it around a spool. In this work, we discover the nanoscale equivalent to spooling yarn, which can be used to deterministically unravel a significant portion of DNA, or other macromolecules, in a single-file, non-overlapping way around a micron-size cylinder. We call this effect spin-induced macromolecular spooling.

VII. ACKNOWLEDGMENTS

This work was primarily supported by funding from the NSERC Discovery Program, McGill University and the University of Ottawa. T.N.S. would like to acknowledge fellowship funding from EMBO (ALTF181-2013) and an ERC Advanced Grant MiCE (291234).

---

[1] J. C. Hansen, The EMBO Journal **31**, 1621 (2012), ISSN 0261-4189.

[2] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struc. Biol. **10**, 279 (2000), ISSN 0959-440X.

[3] C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003).

[4] C. B. Renner and P. Doyle, Soft Matter (2015).

[5] R. M. Jendrejack, D. C. Schwartz, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **120**, 2513 (2004).

[6] O. B. Usta, J. E. Butler, and A. J. C. Ladd, Phys. Fluids **18** (2006).

[7] K. Jo, Y.-L. Chen, J. J. de Pablo, and D. C. Schwartz, Lab Chip **9**, 2348 (2009).

[8] F. Persson and J. O. Tegenfeldt, Chem. Soc. Rev. **39**, 985 (2010).

[9] D. J. Berard, F. Michaud, S. Mahshid, M. J. Ahamed, C. M. J. McFaul, J. S. Leith, P. Bérubé, R. Sladek, W. Reisner, and S. R. Leslie, PNAS **111**, 13295 (2014).

[10] M. Fyta, J. Phys. Condens. Matter **27**, 273101 (2015).

[11] A. Cournac, R. Koszul, and J. Mozziconacci, Nucleic Acids Research **44**, 245 (2016).

[12] T. T. Perkins, D. E. Smith, and S. Chu, Science **276**, 2016 (1997).

[13] T. Saito, T. Sakaue, D. Kaneko, M. Washizu, and H. Oana, J. Chem. Phys. **135**, 154901 (2011).

[14] B. Ladoux and P. S. Doyle, EPL (Europhysics Letters) **52**, 511 (2000).

[15] Y. Gratton and G. W. Slater, The European Physical Journal E **17**, 455 (2005), ISSN 1292-8941.

[16] M. D. Graham, Annu. Rev. Fluid Mech. **43**, 273 (2011).

[17] K. Günther, M. Mertig, and R. Seidel, Nucleic Acids Res **38**, 6526 (2010), ISSN 0305-1048.

[18] A. Balducci, P. Mao, J. Han, and P. S. Doyle, Macromolecules **39**, 6273 (2006).

[19] D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman, Macromolecules **46**, 8369 (2013).

[20] A. C. Klepinger, M. K. Greenier, and S. L. Levy, Macromolecules **48**, 9007 (2015).

[21] R. E. Teixeira, H. P. Babcock, E. S. G. Shaqfeh, and S. Chu, Macromolecules **38**, 581 (2005).

[22] C. M. Schroeder, R. E. Teixeira, E. S. G. Shaqfeh, and S. Chu, Phys. Rev. Lett. **95**, 018301 (2005).

[23] G. W. Slater, C. Holm, M. V. Chubynsky, H. W. de Haan, A. Dubé, K. Grass, O. A. Hickey, C. Kingsburry, D. Sean, T. N. Shendruk, et al., Electrophoresis **30**, 792 (2009).

[24] W. Jiang, J. Huang, Y. Wang, and M. Laradji, J. Chem. Phys. **126**, 044901 (2007).

[25] C. M. Schroeder, R. E. Teixeira, E. S. G. Shaqfeh, , and S. Chu, Macromolecules **38**, 1967 (2005).

[26] M. Harasim, B. Wunderlich, O. Peleg, M. Kröger, and A. R. Bausch, Phys. Rev. Lett. **110**, 108302 (2013).

[27] F. Brochard-Wyart, EPL (Europhysics Letters) **23**, 105 (1993).

[28] M. Laleman, M. Baiesi, B. P. Belotserkovskii, T. Sakaue, J.-C. Walter, and E. Carlon, Macromolecules **49**, 405 (2016).

[29] G. W. Slater, Y. Gratton, M. Kenward, L. McCormick, and F. Tessier, Soft Materials **2**, 155 (2004).

[30] A. Buguin and F. Brochard-Wyart, Macromolecules **29**, 4937 (1996).

F

APPENDIX: Physical confinement signals regulate the organization of stem cells in three dimensions

Sebastian V. Hadjiantoniou, **David Sean**, Maxime Ignacio, Michel Godin, Gary W. Slater and Andrew E. Pelling

Published in *Journal of the Royal Society Interface*
Sebastian V. Hadjiantoniou, David Sean, Maxime Ignacio, Michel Godin, Gary W. Slater and Andrew E. Pelling: Journal of the Royal Society Interface. 2016, **13**. 20160613

Permission to reproduce from Royal Society Interface Publishing was not granted.
The following Appendix contains the manuscript version of the paper. The published version can be found online at the following URL: http://doi.org/10.1098/rsif.2016.0613.

Physical Confinement Signals Regulate the Organization of Stem Cells in Three Dimensions

Sebastian V. Hadjiantoniou[1], David Sean[2], Maxime Ignacio[2], Michel Godin,[2,3,4], Gary W. Slater[2] and Andrew E. Pelling[1,2,6]

[1]*Department of Biology, Gendron Hall, 30 Marie Curie, University of Ottawa, Ottawa, ON, K1N 6N5, Canada*
[2]*Department of Physics, MacDonald Hall, 150 Louis Pasteur, University of Ottawa, Ottawa, ON, K1N 6N5, Canada*
[3]*Department of Mechanical Engineering, Site Building, 800 King Edward Avenue, University of Ottawa, Ottawa, ON, K1N 6N5, Canada*
[4]*Ottawa-Carleton Institute for Biomedical Engineering, Ottawa, Ontario, K1N 6N5, Canada*
[6]*Institute for Science, Society and Policy, Desmarais Building, 55 Laurier Ave. East, University of Ottawa, Ottawa, ON, K1N 6N5, Canada*

Keywords: Biophysics, Cell biology, Biotechnology, Stem cells, Molecular Biology

Author for correspondence: a@pellinglab.net.

ABSTRACT

During embryogenesis, the spherical inner cell mass (ICM) proliferates in the confined environment of a blastocyst. Embryonic stem cells (ESCs) are derived from the ICM, and mimicking embryogenesis *in vitro*, mouse ESCs (mESCs) are often cultured in hanging droplets. This promotes the formation of a spheroid as the cells sediment and aggregate due to increased physical confinement and cell-cell interactions. In contrast, mESCs form 2D monolayers on flat substrates and it remains unclear if the difference in organization is due to a lack of physical confinement or increased cell-substrate vs cell-cell interactions. Employing microfabricated substrates we demonstrate that a single geometric degree of physical confinement on a surface can also initiate spherogenesis. Experiment and computation reveal that a balance between cell-cell and cell-substrate interactions finely controls the morphology and organization of mESC aggregates. Physical confinement is thus an important regulatory cue in the 3D organization and morphogenesis of developing cells.

It is now well recognized that physical cues play an important role in the differentiation and fate of stem cells(1–4). Mechanical forces, matrix topography, matrix mechanics and even cell shape have a profound influence on lineage commitment(1,3–5). However, occurring well before commitment, the inner cell mass (ICM) must organize into a 3D spherical aggregate inside of the blastocyst, confined between a fluid filled cavity and the outer trophoblast membrane. These conditions permit the initiation of embryogenesis, whereby mESCs will subdivide into three germ layers and go on to differentiate into every cell type in the body.

It has been well established that during their pluripotent state, mESCs express high levels of cell adhesion proteins, most prominently E-cadherins(6–8). This is unsurprising as preliminary formation of the ICM is heavily dependent on cell-cell interactions which dynamically form the physical properties of their microenvironment. In turn, E-cadherin expression and regulation play a pivotal role in morphogenesis and development through biomechanical feedback events(9,10). The spatiotemporal pattern of expression of E-cadherins is responsible for cell layer separation, recognition and rearrangement(11). Furthermore, there is also an intimate association between intercellular adhesion sites, the actomyosin network and associated regulatory pathways(12,13). It has been reported that cortical actin is responsible for promoting distinct patterns of cadherin migration and clustering along the rim of the contact plane between cells(14,15). Clearly, there exist important regulatory mechanisms to control the adhesion and organization of mESCs during the early formation of the ICM.

In vitro, standard protocols have been devised to produce cell aggregates defined as embryoid bodies (EBs) in an effort to mimic the processes that give rise to the ICM. Widespread methodologies attempt to replicate the spherical aggregation and proliferation of the ICM by confining mESCs in a hanging drop, which forces cell-cell aggregation due to the lack of a substrate. In addition to hang drop cultures, microfabricated substrates containing spatially ordered arrays of semi-spherical wells have also been utilized to initiate the aggregation of mESCs into a spheroidal structure(16–18). Cells seeded onto such substrates are passively confined within the wells that act as a mold for the developing EB. Microfabricated substrates have the advantage of allowing one to exert control over the size and geometry of the resulting EB, which ultimately influences downstream differentiation(16,17,19).

It is thought that semi-spherical microwells are simply acting as a geometric template for the resulting spherical EB. However, it is possible that the spherical formation of the EB is also finely controlled through distinct biochemical and biophysical regulatory mechanisms. Therefore, we hypothesized that spherical EBs would continue to form spontaneously even in the presence of anisotropic, non-circular physical confinement. Moreover, such aggregation and organization should be driven by both physical and biochemical mechanisms. Contrary to other studies that utilize isotropic microwells(16), we've chosen to systematically examine this issue by fabricating substrates containing 1.5cm long, 100 μm deep open-top grooves of varying widths (50-1000 μm)(20). On these substrates, cells quickly accumulate in the bottom of the grooves, as they possess a higher mass density than the surrounding medium. After accumulating in the bottom and adhering, cells are free to migrate out of the grooves and along their length. However cell migration is limited across the groove width. We demonstrate that this single axis of confinement (arising from the groove width) is all that is required to induce the spherical and isotropic aggregation of mESCs into an early EB. During the first ~6hrs after entering the grooves, physical confinement leads to an increase in cell-cell collisions which nucleate the formation of a spherical EB. Conversely, in the absence of any physical confinement (cells on planar surfaces) mESCs tended to form large, flat islands, even when seeded at high density.

Finally, physical confinement alone is not enough to induce EB formation, as E-cadherin activity and actomyosin dynamics play a key regulatory role. The cytoskeleton constitutively exerts tension on E-cadherins at the plasma membrane capable of modulating E-cadherin expression and subsequent cell-cell binding.(21–23) Therefore, the molecular mechanisms that control cell-cell adhesion and aggregation are also important during the earliest stages of spherical EB formation. Although mechanical forces and material properties are clearly important during developmental processes (1,24), our work highlights the fact that physical confinement also plays an important role in early embryogenesis. One of the earliest stages of lineage specification occurs in a highly confined environment that itself imparts a regulatory role in the aggregation and organization of cells into the ICM. Our work suggests this physical separation of the ICM and the trophoblast is not only biochemically regulated, but also driven by physical signalling.

2. Materials and Methods

2.1 Substrate fabrication

Master substrates were fabricated by means of standard soft photolithography techniques on polished silicon wafers (Unicersitywafers.com, USA). Impurities on the wafer were removed with a Piranha solution (3:1 sulphuric acid: hydrogen peroxide) followed by a rinsing in de-ionized water and finally baked at 200°C for 30 mins. A 100 µm uniform thickness of SU-8 2050 photoresists (Microchip, USA) was then spin-coated onto the wafer. Photomask patterns were subsequently transferred to the photoresist via UV exposure for 10 sec as per the manufacturer protocol. The photomask consisted of separate 2.25 cm^2 square regions each containing 1.5 cm long black lines, with 100 µm spacing. The widths of the lines varied among square regions from 50, 100, 200, 500 to 1000 µm. Polydimethylsiloxane (PDMS) moulds with channeled topographies were formed by pouring a 1:10 (curing agent: elastomer) (Sylgard 184, Ellsworth Adhesives) over the photoresist master mold. The PDMS was allowed to crosslink in a convection oven at 80°C for 3 hrs. Functionalization of the PDMS to permit collagen coating was achieved through air plasma treatment at 50 w for 30 sec. Rat-tail collagen I (5 $\mu g/cm^2$, Gibco) was then deposited onto the PDMS at room temperature for 30 mins to allow adhesion followed by a phosphate buffered saline (PBS) rinse.

2.2 Cell culture and drug studies

D3 mouse embryonic stem cells (mESC) (ATTC, #CRL-1934) were cultured as per ATTC's protocol(25), in Dulbeco's modified Eagle's medium (DMEM) supplemented with 12.5% Fetal Bovine Serum, 0.1 mM non-essential amino acids, 30 μg/ml Gentamicin (Gibco), 0.1 mM 2-mercaptoenthanol (Sigma) and 1000 U/ml leukemia inhibitory Factor (LIF) (Millipore, Catalogue # ESG1106). Cells were passaged every 48 hrs to prevent differentiation. Cells were seeded onto the PDMS topographies immediately after passaging at 50 000 cells/dish. Inhibition studies of Rho-kinase (Y-27632; 10 μM, Sigma, Catalogue #Y0503), Myo II (Blebbistatin; 10 μM, Sigma, Catalogue #B0560) and mDia (SMIFH2; 10 μM, Sigma, Catalogue #S4826) were all performed by exposing mESC for the 48 hrs incubation time period.

2.3 E-cadherin blocking

Prior to seeding, monoclonal DECMA-1 anti-uromodulin/e-cadherin antibody (Sigma; 1:1600, Catalogue # U3254) was added to cells in free suspension for 30 min. This antibody binds directly to uvomorulin/E-Cadherin which has been characterized as a 120 kDa cell surface glycoprotein. Due to their protein structure similarities, the antibody also binds to L-CAM and Cell CAM 80/120, additional transmembrane proteins involved in cell adhesion. Cells were then centrifuged and re-suspended in fresh media and plated as described in section 2.2. Verification of primary antibody binding was performed with an anti-rat IgG CF488A secondary antibody (Sigma, 1:500) during immunofluorescent staining (26,27).

2.4 Immunofluorescence staining, quantification, time lapse imaging and microscopy

Cells cultured on PDMS substrates were fixed with 3.5% paraformaldehyde and permeabilized with Triton X-100 at 37 °C. Cells were stained for: vinculin, using monoclonal anti Vinculin (Sigma, Catalogue # V9131), actin, using phalloidin conjugated to Alexa Fluor 546 (Invitrogen, Catalogue #A22283) and DNA using DAPI (Invitrogen, Catalogue #D1306). A full protocol has been published previously (28). Samples were then mounted using Vectashield (Vector Labs) and a #1 coverslip placed on top of the PDMS substrate. Samples were then inverted and imaged with a Nikon Ti-E A1-R high-speed resonant laser scanning confocal microscope (LSCM) with a phase contrast 10x NA0.3 objective or a DIC 60x NA1.2 water immersion objective.

Immunofluorescence quantification of vinculin was performed by uniformly staining all samples concurrently and image capturing at set parameters. Images were then processed in Image J whereby the integrated density is relatively compared to background fluorescence using the measure tool. (29) For time-lapse imaging, cells were seeded onto the PDMS substrate with an additional 3 ml of media. Time-lapse phase contrast imaging was carried out on an AE30 Motic microscope with a 10x objective, which was enclosed in a temperature regulated (37°C) box. Samples were imaged every 10 mins for up to 48 hrs. Scanning Electron Microscope images were acquired on a JEOL JSM-7500F FESEM.

2.5 Image and statistical analysis

To obtain a quantitative description of the changes in aggregate morphology, we developed two metrics defined as planar isotropy (I_p) and globular isotropy (I_g). The planar isotropy describes the circularity of the aggregate in the *xy* plane, whereas globular isotropy describes its sphericity by comparing its vertical growth (*z*-direction) to its effective *xy* size. A three-dimensional mesh representation of the cell surface is generated in ImageJ (30) from the confocal image slices using a marching cube algorithm(31). The mesh vertices $\vec{r} = (r_x, r_y, r_z)$ are used to define the object's position $(\langle r_x \rangle, \langle r_y \rangle, \langle r_z \rangle)$, where $\langle\ \rangle$ denotes an unweighted mean over the vertices. The root mean square distance of all the points to this center position is then used as a metric describing the extent of the aggregate in the three Cartesian directions. Thus $R_x = \sqrt{\sigma_{xx}}$, $R_y = \sqrt{\sigma_{yy}}$ and $R_z = \sqrt{\sigma_{zz}}$ where

$$(1)\ \sigma_{ij} = \langle\ (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle)\ \rangle,$$

for $i, j = \{x, y, z\}$. To calculate the planar isotropy, we first project the points onto the bottom surface (the *xy*-plane) and calculate the principal components of the resulting data set. This requires finding the eigenvalues and eigenvectors of the covariance matrix

$$(2)\ \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}.$$

The ratio between the effective length along the major ($R_{max} = \sqrt{\sigma_{max}}$) and minor ($R_{min} = \sqrt{\sigma_{min}}$) axes (from the eigenvalues σ_{max} and σ_{min} respectively) then defines the dimensionless parameter

$$(3)\ planar\ isotropy = I_p = \frac{R_{min}}{R_{max}}.$$

Note that a value of $I_p \approx 1$ is expected for an isotropic sample (with a roughly circular footprint) while $I_p < 1$ is indicative of an elongated profile.

In addition, we define the dimensionless *globular isotropy* as a measure of the effective height R_z relative to its effective planar size $R_p = \sqrt{\frac{1}{2}({R_{max}}^2 + {R_{min}}^2)}$. This is calculated using the formula

$$(4)\ \ globular\ isotropy = I_g = \frac{R_z}{R_p} = \sqrt{\frac{2\sigma_{zz}}{(\lambda_{max} + \lambda_{min})}}.$$

With this definition, we expect $I_g < 1$ for the usual 'flat' (two-dimensional growth) profile and $I_g \approx 1$ for a spherical (three-dimensional growth) profile.

2.6 Statistics

All statistical analyses were performed using a one-way ANOVA followed by a Tukey test for mean comparison. Unless otherwise stated, all data is presented as mean ± s.e.m. Each condition, consisting of the various drugs and channel widths were duplicated 3 times.

2.7 Simulations

In order to elucidate the dependence of the cluster morphology upon both geometrical confinement and cell-cell/cell-substrate interactions, a simple simulation model is used where these factors can be independently controlled. Additional factors which can possibly influence morphology, such as cell interaction range, initial cell surface density, and initial cell seed amount are held constant. This simulation model is used as a tool to reveal the potential influencing physical factors observed in aggregate formation and does not attempt to fully represent the complexities of dynamic biological systems.

We thus use coarse-grained Langevin-Dynamics simulations where cells are described as single spherical beads. Individual cells are subject to forces arising from gravity, the solvent, the substrate, as well as other cells in the system. The equation of motion for the simulation beads is given by the Langevin equation (32)

$$m\frac{d^2\boldsymbol{r}_i(t)}{dt^2} = -\nabla V + \boldsymbol{F}_{\text{grav}} + \boldsymbol{F}_f + \boldsymbol{F}_B \ (2)$$

where m is the mass of the cells, $\boldsymbol{r}_i$ is the position of the i^{th} cell, V is the net interaction potential, and $\boldsymbol{F}_{\text{grav}}$ is the gravitational force. The last two terms are used to implicitly model the solvent as contributing to a dissipative friction $\boldsymbol{F}_f$, and an effective Brownian force $\boldsymbol{F}_B$ (32)

For simplicity, the cell-cell and cell-substrate interaction potentials have an identical form so that they can be controlled via a single parameter. This is achieved using the modified Lennard-Jones potential:

$$V_{\text{LJ}} = \begin{cases} -F_{\text{cap}} r + A_1 & \text{if } r < r_{\text{cap}}, \\ 4\,\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) + A_2 & \text{if } r_{\text{cap}} < r < r_{\text{cut}}, \\ 0 & \text{if } r_{\text{cut}} < r, \end{cases} \tag{1}$$

with r being the distance between a cell and an object (either another cell or a substrate surface), ε is the depth of the potential well, and σ is the effective size of the cell (see Supplementary Fig.1). First, for short distances ($r < r_{\text{cap}}$) we define a maximum repulsive force of F_{cap}. This force cap is needed due to our chosen duplication method that places overlapping daughter cells unto mother cells. The constant A_1 is chosen such that the potential is continuous at $r = r_{\text{cap}}$. The Lennard-Jones potential is used for intermediate distances $\left(r_{\text{cap}} < r < r_{\text{cut}}\right)$ and has a well minimum at $r_{\text{m}} = 2^{1/6}\sigma$. The A_2 offset is chosen such that the potential is continuous at r_{cut}. We use different well-depths ε_{CC} and ε_{CS} for the cell-cell and cell-substrate interactions, respectively. A finite interaction range is enforced by using a cut-off distance of $r_{\text{cut}} = 2r_{\text{m}}$. Using these three components in the interaction potential maintains a repulsive behavior at short distances and an attractive component over a finite distance (see Supplementary Fig.1).

A single channel is constructed using three mathematical surfaces placed as a bottom plate at $z=0$ surrounded by two walls positioned at $y = \pm w/2$, where w is the channel width. Periodic boundary conditions (Supplementary Fig.1) are used in the x-direction with the channel nominal length l_x chosen such that we achieve a constant cell number density $\rho = N_{\text{init}}/wl_x$ (to match a selected experimental value C= 450 *cell/mm²)* for all widths. This implies that with an initial seed of $N_{\text{init}} = 45$ cells, the simulation system has an area corresponding to 100 μm^2. Under this construction, all the simulation results are taken from systems that have the same cell count and the same cell density despite having varying channel widths. Thus, by only changing the channel aspect-ratio (while keeping the area of the channel floor the same), these simulations allows us to de-couple growth effects arising from different cell numbers and densities inside the channels.

The simulation is structured into two distinct phases, i) initial cell diffusion; and ii) cell duplication. After the cells are initialized randomly unto the bottom surface, they are allowed to diffuse and explore the channel during an initial period of time (10^5 timesteps), during which they are allowed to coalesce into small clusters and / or fall into low-energy regions such as along the edges of the channels. Since cell diffusion is observed to be supressed after approximately 6 hours due to integrin binding, the duration of this phase was chosen when the bead configurations were found to be similar to those obtained from the experimental images of freshly attached cells.

At the end of this diffusion stage, daughter cells are added into the simulation until a total population of $N_{\mathrm{tot}} = 352$ cells is reached. This corresponds to doubling the initial population of N_{init}=44 cells three times (which is thus equivalent to 45 hours of incubation time). Duplication is conducted by selecting a random mother cell from the simulation and inserting an overlapping daughter cell at the same location. We find that a short lapse is sufficient to allow the mother-daughter cell duo to relax and move away from one another. During this short time period, the cells are still allowed to diffuse. This sequence is repeated until all potential mother cells have duplicated once, which marks the end of a doubling phase.

The described simulation is implemented with the ESPResSO package (33) and visualized with VMD (34). Supplementary Table 1 contains the ESPResSO numerical values for the parameters described herein. We find that these values yield stable integration in the over-damped limit and provide simulated trajectories that are qualitatively in agreement with those observed in the laboratory. Given the generic nature of this model, we nevertheless find the dependence of the cluster morphology upon the channel width (Fig 4). A systematic study of these parameters remains to be conducted, as it is out of the scope of the current study.

3. RESULTS

3.1 Physical confinement promotes the spontaneous formation of 3D spheroids

Standard soft lithography techniques were employed to fabricate collagen coated PDMS substrates containing microfabricated grooves. Groove width was systematically varied (50, 100, 200, 500, 1000 μm) in order to alter the degree of physical confinement on scales 1-2 orders of magnitude larger than the average length of an individual cell (10 μm). Importantly, such geometries act to confine cell movement across the groove width, yet permit movement along the length and out of the groove(20). We have previously shown that this can have profound impacts on the organization and migration characteristics of epithelial and fibroblast cells, even in co-culture(20). In the current study, SEM and phase contrast imaging 48 hrs after plating reveals that the vast majority of mESCs were found to have spontaneously formed spherical aggregates resembling EBs (Fig. 1b, c)(Supplementary Video 1), including many that do not touch either vertical wall. Samples were also fixed and fluorescently labelled for actin and DNA, followed by imaging with laser scanning confocal microscopy (LSCM). Images reveal that physical confinement in the grooves clearly results in mESC aggregates that possess a 3D morphology (Fig. 1d, e). On flat, collagen coated PDMS substrates, mESCs have essentially formed flat islands (1-2 cells thick), rather than a structure resembling a spheroid (Fig. 1f, g).

To quantify the morphology of the mESC aggregates observed in this study, we calculated their planar (I_p) and globular (I_g) isotropy (Materials and Methods). After mESCs were allowed to proliferate on flat or grooved substrates for 48 hrs, they were fixed and stained as described above and imaged with LSCM. For each substrate, 10 randomly chosen locations were imaged on both substrates. For each aggregate identified, the images were thresholded and for each image slice in the confocal stack, we determined the coordinates of the circumference of the aggregate using the 'analyze particles' ImageJ plugin. This approach provided the coordinates of the aggregates in 3D and allowed us to quantitatively assess the shape of the aggregates using our definitions of I_p and I_g. As the isotropy value approaches 1, I_p and I_g describe a perfect circle and sphere respectively. As expected, when one views the aggregates from above, regardless of substrate topography, all aggregates appear roughly circular with an I_p that varies between 0.66 ± 0.03 and 0.78 ± 0.03 depending on grove width. Importantly, there was no statistically significant dependence of I_p on groove width or substrate topography ($p > 0.05$ in all cases).

In contrast, I_g analysis revealed that on flat substrates, cells proliferated in a planar fashion where $I_g = 0.27 \pm 0.02$. Conversely, in the grooves, globular isotropy was inversely proportional to groove width. In the 50 μm wide grooves, spheroids possessed an average I_g of 0.86 ± 0.03, which decreased to 0.32 ± 0.02 in the 1000 μm wide grooves (Fig. 1h). Only the aggregates in the 50 and 100 μm grooves possessed an I_g significantly larger than the aggregates on the flat substrates ($p<0.001$). Interestingly, the number of cells per aggregate (50 ± 7 cells) did not display any statistically significant dependence on groove width (Fig. 1i).

An SEM image of a spheroid in a 100 μm channel after 48 hrs of proliferation (Fig. 1b) clearly reveals that spheroids can form without contacting channel walls for support, consistent with LSCM data. However, spheroids can also be found in contact with groove walls if they become very large or happen to initiate immediately beside a wall. This presented a problem in the 50 μm channels as the aggregates were often found in contact with both walls, potentially leading to the larger observed I_g values. Therefore, in the following sections, we performed all experiments in 100 μm wide channels in order to ensure that the observed morphology is not influenced by the aggregate coming into physical contact with both channel walls.

3.2 The role of actin dynamics on 3D aggregate formation

In order to investigate the role of actin dynamics in regulating aggregation and 3D morphology, mESCs were treated with a well-known Rho-kinase (ROCK) inhibitor (Y27632), a specific myosin-II (myoII) inhibitor (blebbistatin) and an mDia1/2 inhibitor (SMIFH2). In general, ROCK plays a major role in many aspects of actin organization, myo-II governs actin contractility and mDia1/2 belong to a family of formins that regulate actin nucleation and polymerization(35–37). After 48 hrs of culture, cells were fixed and stained to visualize the actin cytoskeleton and the nucleus with LSCM. Inhibiting ROCK and mDia1 caused a statistically significant effect ($p<0.001$) on spheroid shape characteristics whilst blebbistatin had little effect. The ROCK inhibitor completely inhibited the 3D shape of the spheroid (Fig. 2a), resulting in cells growing along the sides of the grooves. Cells could be found in isolation and small island-like aggregates that resembled growth on a flat substrate. The average I_g of the aggregates was determined to be 0.22±0.02, which is significantly less than untreated cells 0.49±0.04 ($p<0.001$). Blebbistatin treatments appeared to have no effect on I_g (0.54±0.05) compared to untreated cells in 100 μm channels (Fig. 2b). Conversely, inhibition of mDia resulted in the formation of more isotropic spheroids ($I_g = 0.80\pm0.04$, Fig. 2c), significantly higher than the untreated cells in 100 μm channels ($p<0.001$) (Fig. 2d). Finally, on flat substrates, drug treatments had no statistically significant effect ($p>0.05$) on I_p or I_g compared to untreated cultures (Supplementary Fig. 2). These results clearly demonstrate that the role of actomyosin dynamics in regulating cellular aggregation and organization in 3D spheroids is strongly dependent on the presence of physical confinement.

3.3 Direct modification of cell-cell and cell-substrate adhesion

To investigate the importance of cell-cell and cell-substrate interactions during aggregate formation we designed two additional experiments. In the first case, we interfered with cell-cell interactions by treating mESCs in suspension with an E-cadherin primary antibody for 30 mins prior to culturing on flat and 100 μm grooved surfaces. After 48 hrs of culture, cells were stained and imaged for actin and DNA. As well, we also treated the cells with a fluorescently labelled secondary antibody to visualize the E-cadherin antibody that was introduced prior to plating. Applying the secondary antibody confirmed the presence of the primary antibody bound to E-cadherins even after 48 hrs in culture. Imaging reveals that cell-cell aggregate formation was significantly impaired. Individual cells were clearly dispersed and sometimes found in isolation (Fig. 2e). Spheroid formation was abolished and cells were found throughout the channels.

In a second experiment we sought to manipulate cell-substrate adhesion. In this case, PDMS substrates were either left unfunctionalized or functionalized with fibronectin in addition to collagen. Bare PDMS channels or flat substrates completely lacked any significant cell proliferation after 48 hrs confirming the importance of matrix proteins in promoting adhesion (Supplementary Fig. 3a, b). We then hypothesized that additional extracellular matrix proteins should provide an increased number of sites for cell-substrate binding via integrins and thereby enhance cell-substrate interactions. Interestingly, the altered adhesive properties of the substrate (collage + fibronectin) did not impede the formation of spheroids, which occurred in a manner consistent with previous data. However, the addition of fibronectin to the surface promoted cell growth on the groove ridges (Supplementary Fig. 3c). The morphology of the cells on the ridges was similar to cells growing on flat substrates. Cells found on ridges clearly lack the influence of physical confinement, indicating its importance in the formation of spheroids.

3.4 Effects of confinement on cell displacement and collision frequency

Thus far, we have reported on the appearance of spheroids in confined geometries after 48 hrs of culture. In order to understand the initiation and progression of spheroid formation at early times we performed live cell time-lapse phase contrast imaging of mESCs over a 24 hour period following plating on flat and grooved surfaces (3 replicates in both cases). On flat substrates (Fig. 3a)(Supplementary Video 1), cells initially attach to the surface as single cells or in small groups and then proliferate into island shaped aggregates. In general we also observe that island growth occurs mainly through proliferation rather than through new cells joining the island through migration.

Time lapse imaging of grooved surfaces reveals a very different series of events. Immediately after seeding, cells were found at the bottom of the grooves due to their higher mass density than the surrounding medium (Fig. 3b). Within 12hrs incubation time, cell-cell aggregation nucleates the formation of the spheroids. By 24hrs, the spheroids have exhibited proliferative growth, expanding in 3D (Supplementary video 2). A more in depth analysis over a two hour period shortly after seeding (Fig. 3c)(Supplementary Video 3), revealed that very early spheroid formation is driven by the collision and aggregation of a small number of cells. In the early spheroid, mitosis can be observed (Fig. 3d). After cytokinesis occurs the newly divided cells remain adhered to the aggregate. This process continues as the initial nucleate develops into a 3D spheroid.

To quantify these observations, we measured the frequency of collision events between cells during the early formation of the spheroids (Fig. 4a). Importantly, cells found in the grooves exhibited a collision frequency (22 ± 7 collisions/hr) ~4-fold higher than cells on flat surfaces (7 ± 2 collisions/hr). However, as the number of cells seeded onto each substrate was constant, the collection of cells into the grooves leads to a higher effective cell density as a result of physical confinement. To control for this effect, a series of experiments were performed in which the number of cells added to the flat substrate matched the effective density observed in the 100 μm channels (~450 cells/mm^2). Importantly, even at this higher seeding density the observed collision frequency (3 ± 1 collisions/hr) remained significantly lower than observed within the 100 μm channels ($p<0.001$). Interestingly, when comparing the effect of cell density on flat substrates, there appears to be a two fold drop in cell collision frequency within high density conditions. Statistical analysis shows no significant difference ($p<0.05$) between these conditions which suggests that the discrepancy is attributed to the inherent variability between experiments. Although islands were still observed to form at higher seeding density, they tended to be much larger, as expected (Supplementary Fig. 3d,f). In our initial experiments, cell aggregates in 500 μm channels were geometrically indistinguishable from aggregates on flat substrates (Fig. 1h). Therefore, we also conducted an experiment where the higher seeding density was applied to the 500 μm channels as well. In this case, we observed a significantly lower collision frequency (7 ± 4 collisions/hr) than in the 100 μm channels ($p<0.001$). This suggests that increased cell density alone is not responsible for the observed changes in collision frequency and that physical confinement is playing an influential role in promoting cell-cell interactions.

3.5 Simulation of Spheroid Formation

Depending on their respective adhesion affinities, cell-cell and cell-substrate interactions can significantly alter cell morphology and behaviour(7,38–40). To assess the relative importance of cell-cell and cell-substrate binding in spheroid formation, we developed a simulation model that recreates the confinement conditions of our experiments (Fig.4b). It is important to note that this model does not reproduce all the biological complexities of the system but presents complementary information which provides insight into the physical interaction between cell and substrate. In either 100 μm channels or flat substrates, cells are allowed to diffuse, interact with each other and the substrate, and undergo mitosis. Through the utilization of a coarse-grain Langevin Dynamic model, we altered the binding energies between cells ($\in_{cc}$) and substrate ($\in_{cs}$) and quantitatively analyzed the effect on the development of spheroid formations. The relative change in selective binding is denoted by chi ($\chi = \in_{cs}/\in_{cc}$), whereby $\in_{cc}$ is kept constant and $\in_{cs}$ is increased. As the value of χ increases, cell-substrate attraction becomes stronger. With an initial cell density of 450 cells/mm^2, the simulation initiates the diffusion phase, whereby cells explore the surrounding space. Random cells are then chosen periodically to duplicate followed by a relaxation phase whereby newly formed cells can move to a lower energy position. This sequence is repeated until each cell has undergone three generations of duplications.

As can be expected, under extreme χ values of 0 and 4, cell aggregation was extremely spheroidal or flat respectively (Supplementary Fig. 1). Interestingly however, changing the channel width within any condition of χ displayed significant ($p<0.001$) changes in the I_g of the aggregates. At $\chi = 1$, the simulation represents similar I_p and I_g values acquired experimentally. In a non-confined system (flat), the simulation displayed aggregates with an average I_g of 0.39±0.005, negligibly higher ($p<0.01$) than the value of 0.27±0.02 acquired experimentally. Similarly, under the 100 μm confinement condition, the simulated and experimental I_g values were 0.46±0.005 and 0.55±0.05 respectively. These results suggest that a relatively balanced $\in_{cc}$ and $\in_{cs}$ is required to induce spherogenesis.

4. Discussion

In this study we examined the influence of physical confinement on the three-dimensional spatial organization and growth of mESCs. To investigate this, we fabricated grooved, collagen-functionalized, PDMS substrates of varying widths (50, 100, 200, 500, 1000 μm) and constant depth (100 μm). Interestingly, after 48 hrs of incubation, cells grown on substrates with 50-200 μm grooves displayed clear spheroidal growth. This differs from the traditional flat two-dimensional cultures in which cells display flat, island-shaped aggregates. Importantly, in the absence of collagen-functionalization, cells did not adhere or proliferate on the PDMS substrates. Contrary to previous strategies utilizing concave microwells(16) to form spheroids, this experiment reveals that spontaneous 3D growth can occur as a result of a single axis of physical confinement. Quantitative globular isotropy analysis revealed how 3D growth decreased with decreasing physical confinement (Fig. 1h). It is interesting to note that the I_p within flat conditions was <0.8, suggesting an intrinsic level of anisotropy. Although speculative, this intrinsic level of anisotropy may potentially arise from actin stress fibre polarization in individual cells in the aggregate. Furthermore, spheroid formation did not appear to be dependent on the proximity to channel walls, displaying 3D growth without a surrounding matrix. The aggregation was also not the result of a higher cell density, as there was no statistical difference ($p>0.05$) in the collision frequency of highly seeded (450 cells/mm^2) flat PDMS surfaces.

To examine whether increased substrate adhesion could alter spheroid formation in the channel, we increased cell-substrate interaction by depositing both collagen and fibronectin before cellular plating. Within the confines of the grooves, spheroid formation was not obviously altered**.** Interestingly, cells adhered and grew at the top of the ridges as well, in a flat island like shape, a completely different morphology to the cells below. This occurrence, in which all sets of conditions are the same, isolates confinement as the influencing factor inducing three dimensional growth.

In our previous work, mouse NIH3T3 fibroblasts and Madin-Darby Canine Kidney (MDCK) epithelial cells grown in similar channelled conditions favoured substrate adhesion and eventual monolayer formation(20). The preferential affinity displayed by in vitro mESCs for globular formation vs substrate adhesion mimics in vivo development at the blastocyst stage. Confined within the inner cell mass, E-Cadherins play a pivotal role in cell rearrangement, tissue morphogenesis, establishing cell polarity and tissue architecture maintenance(41,42). It is also highly associated with the actomysin network, as they possess a regulatory feedback loop which can modulate cadherin expression during embryogenesis(7,9,15,43). Expression of E-Cadherin significantly lowers immediately after differentiation, which has made it a pluripotency marker for undifferentiated cells(42,44,45). To examine its influence on spheroid formation within a confined channel, cells were pre-loaded with a primary E-Cadherin antibody to block cell-cell adhesion which caused the abolition of spheroid formation (Fig. 2e). This observation is consistent with the notion that E-cadherin operates within a positive feedback loop, that if perturbed, disrupts colony formation and impairs long-term survival of ESCs(46).

ROCK and myo-II are also known to play a role in regulating the stability of newly formed cell-cell junctions(47,48). Here, ROCK inhibition disrupted spheroid formation, producing results consistent with the E-Cadherin blocking experiments. This is also consistent with our understanding of actin dynamics as inhibiting Rho subsequently effects E-cadherin binding. Surprisingly, myo-II did not have any clear effect on spheroid formation or shape characteristics. In this study we were forced to use a relatively low concentration of blebbistatin (2 μM) as higher concentrations (5 and 10 μM) resulted in the complete loss of all cells from the substrate. Myosin-II inhibition will destabilize E-cadherin cell-cell contacts (47,48) and disrupt cell-substrate contacts (Fig. 5) (49). We speculate there exists a sharp threshold level of myo-II activity that is required to maintain cell-substrate and cell-cell adhesion. Finally, inhibition of formins via SMIFH2 allowed us to directly inhibit actin nucleation, migration and cell-substrate adhesion pathways. As an actin regulator, mDia1/2 accelerates actin polymerization, focal adhesion dynamics, cell-substrate attachment and migration(37,50–52). It mediates this through its interactions with the c-Src pathway and focal adhesion kinase (FAK) pathway. Both c-Src and FAK are key proteins responsible for the formation and dynamic reorganization of focal adhesion complexes(52,53). Formin inhibition resulted in spheroids with significantly larger I_g. As decreased cell-substrate adhesion and inhibited migratory processes tend to promote cell-cell adhesion through cadherin binding, subsequently leading to a more spherical morphology as shown by their globular isotropy values. Finally, in the channels, the number of cell-cell collisions per hour was about 3-fold higher compared to cells on flat substrates. This leads to an increased probability of experiencing a cell-cell collision and adhesion event in the channels.

Taken together, the picture that emerges from the experimental data is that a balance between physical and biochemical factors are influencing the early formation of embryoid bodies. The developmental pathway of the cells depends upon their dynamic interaction with the physical properties that surround them rather than their static position at any moment in development, as suggested by Beloussov et al.(54–56). To explore this idea further, we developed a simulation that could reproduce the organization of cells we observed during experiments. Specific cellular organization could arise from cells allowed to diffuse and collide while undergoing energetically favourable movements during relaxation phases. A parameter χ ($\frac{\epsilon_{cs}}{\epsilon_{cc}}$), was developed to explore this phenomenon by simply changing the relative balance between cell-substrate and cell-cell adhesion strength. At $\chi = 1.0$, whereby the strength of cell-cell adhesion is equal to that of cell-substrate, simulated aggregates resemble the aggregates observed experimentally. Importantly, experimental observations were reproduced under situations in which cell-cell and cell-substrate strength becomes unbalanced. In cases where $\chi < 1.0$, spheroids are formed with much higher I_g than the case where $\chi = 1.0$. This reflects the experimental results obtained with formin inhibition where cell-substrate interaction is impaired significantly. Alternatively, when $\chi > 1.0$, cellular aggregates possessed a low I_g compared to the scenario in which $\chi = 1.0$. In this case, inhibiting cell-cell interactions promotes the likelihood of cell-substrate interactions and the formation of islands, or dispersed groups of cells that do not resemble spheroids.

It is well appreciated that soluble signals (growth factors, cytokines) are highly involved in the self-regulating microenvironment designated the stem cell niche(39,57,58). However, recent evidence also suggests that stem cell development is strongly influenced by coexisting insoluble adhesive, topological and mechanical cues inherently contained in the niche(59–64). Manipulating these physical cues via patterned ECM protein or altered substrate topography has been shown to induce morphological, orientational and proliferative changes in a wide array of cell types(65–69). In this study, we have demonstrated that physical confinement characteristics can also be exploited to control the 3D organization of mESC aggregates. Contrary to flat substrates, mESCs were observed to grow spherically in confined grooves, through initial phases of cellular aggregation followed by proliferative expansion. In recent years, considerable interest has grown in the role physical cues in the microenvironment play in stem cell regulation(2,24,70–72), such as matrix elasticity, nanotopography and stretch. The objective of this work has been to demonstrate that physical confinement also plays an important role in the regulation of stem cell organization in 3D. Therefore, physical confinement can be considered yet another physical cue that stem cells are able to sense and respond to, although its full significance is still being fully elucidated.

Here, we have revealed that topographical confinement can promote 3D spheroidal formation of embryonic stem cells. Cell proliferation in confined space was clearly altered compared to traditional flat 2D cultures. Importantly, confinement is a major factor influencing cell-type dependent response to microtopographies(3,73). This is in line with other recent studies that have manipulated the confinement properties of hydrogels to direct stem cell growth and differentiation(3). Much like the recent elucidation of the importance of substrate stiffness in stem cell fate(24), the evidence of confinement as a critical factor in controlling cell growth is becoming more prominent. With the advent of three dimensional tissue engineering and stem cell regeneration, the phenomenon demonstrated here may further help develop new scaffolds that can direct cell growth and behaviour.

Acknowledgements

This work was supported by individual Natural Sciences and Engineering Research Council (NSERC) Discovery Grants to M.G., G.S. and A.E.P. S.H. was supported by the Queen Elizabeth II Graduate Scholarship in Science and Technology. A.E.P. gratefully acknowledges generous support from the Canada Research Chairs (CRC) program. A.E.P and M.G also acknowledge the support of the Canadian Foundation for Innovation.

Competing Interest

We have no competing interests

References

1. Lee J, Abdeen AA, Zhang D, Kilian KA. Directing stem cell fate on hydrogel substrates by controlling cell geometry, matrix mechanics and adhesion ligand composition. Biomaterials. 2013;34(33):8140–8.
2. Dalby MJ, Gadegaard N, Oreffo ROC. Harnessing nanotopography and integrin-matrix interactions to influence stem cell fate. Nat Mater . Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014 Jun;13(6):558–69.
3. Khetan S, Guvendiren M, Legant WR, Cohen DM, Chen CS, Burdick JA. Degradation-mediated cellular traction directs stem cell fate in covalently crosslinked three-dimensional hydrogels. Nat Mater. Nature Publishing Group; 2013 May;12(5):458–65.
4. Poh Y-C, Chen J, Hong Y, Yi H, Zhang S, Chen J, et al. Generation of organized germ layers from a single mouse embryonic stem cell. Nat Commun. 2014 Jan;5:4000.
5. Dalby MJ, Gadegaard N, Oreffo ROC. Harnessing nanotopography and integrin-matrix interactions to influence stem cell fate. Nat Mater. 2014;13(6):558–69.
6. Soncin F, Ward CM. The function of E-cadherin in stem cell pluripotency and self-renewal. Genes (Basel). 2011;2:229–59.
7. Li L, Bennett SAL, Wang L. Role of E-cadherin and other cell adhesion molecules in survival and differentiation of human pluripotent stem cells. Cell Adh Migr. 2012;6:59–70.
8. Kim Y-S, Yi B-R, Kim N-H, Choi K-C. Role of the epithelial-mesenchymal transition and its effects on embryonic stem cells. Exp Mol Med. Nature Publishing Group; 2014;46(8):1–5.
9. Redmer T, Diecke S, Grigoryan T, Quiroga-Negreira A, Birchmeier W, Besser D. E-

cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep. Nature Publishing Group; 2011 Jul;12(7):720–6.

10. Chou L, Firth JD, Uitto VJ, Brunette DM. Substratum surface topography alters cell shape and regulates fibronectin mRNA level, mRNA stability, secretion and assembly in human fibroblasts. J Cell Sci. 1995 Apr;108 Pt 4:1563–73.

11. Gumbiner BM. Regulation of cadherin-mediated adhesion in morphogenesis. Nat Rev Mol Cell Biol . Nature Publishing Group; 2005 Aug 15;6(8):622–34.

12. Miller CJ, Davidson LA. The interplay between cell signalling and mechanics in developmental processes. Nat Rev Genet . 2013;14:733–44.

13. Fernandez-Gonzalez R, Zallen JA. Cell mechanics and feedback regulation of actomyosin networks. Sci Signal. 2009;2:pe78.

14. de Rooij J. Cadherin adhesion controlled by cortical actin dynamics. Nat Cell Biol . Nature Publishing Group; 2014 May 30;16(6):508–10.

15. Engl W, Arasi B, Yap LL, Thiery JP, Viasnoff V. Actin dynamics modulate mechanosensitive immobilization of E-cadherin at adherens junctions. Nat Cell Biol . Nature Publishing Group; 2014 May 25;16(6):587–94.

16. Choi YY, Chung BG, Lee DH, Khademhosseini A, Kim J-H, Lee S-H. Controlled-size embryoid body formation in concave microwell arrays. Biomaterials . Elsevier Ltd; 2010 May;31(15):4296–303.

17. Hwang Y-S, Chung BG, Ortmann D, Hattori N, Moeller H-C, Khademhosseini A. Microwell-mediated control of embryoid body size regulates embryonic stem cell fate via differential expression of WNT5a and WNT11. Proc Natl Acad Sci U S A . 2009 Oct 6;106(40):16978–83.

18. Karp JM, Yeh J, Eng G, Fukuda J, Blumling J, Suh K-Y, et al. Controlling size, shape and homogeneity of embryoid bodies using poly(ethylene glycol) microwells. Lab Chip . 2007 Jun;7(6):786–94.

19. Discher DE, Mooney DJ, Zandstra PW. Growth factors, matrices, and forces combine and control stem cells. Science (80-). 2009;324:1673–7.

20. Leclerc A, Tremblay D, Hadjiantoniou S, Bukoreshtliev N V, Rogowski JL, Godin M, et al. Three dimensional spatial separation of cells in response to microtopography. Biomaterials . Elsevier Ltd; 2013 Jul 27;34(33):8097–104.

21. Mammoto T, Mammoto A, Ingber DE. Mechanobiology and Developmental Control.. Annual Reviews ; 2013;

22. Strzyz P. Cell adhesion: SUMO controls a tug of war at junctions. Nat Rev Mol Cell Biol . Nature Research; 2015 Oct 22;16(11):641–641.

23. Leckband DE, Rooij J de. Cadherin Adhesion and Mechanotransduction. http://dx.doi.org/101146/annurev-cellbio-100913-013212. Annual Reviews ; 2014;

24. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix Elasticity Directs Stem Cell Lineage Specification. Cell . 2006;126(4):677–89.

25. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. The in vitro development of

blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol . 1985 Jun;87:27–45.

26. Mohamet L, Lea ML, Ward CM. Abrogation of E-cadherin-mediated cellular aggregation allows proliferation of pluripotent mouse embryonic stem cells in shake flask bioreactors. PLoS One . Public Library of Science; 2010 Jan 23;5(9):e12921.
27. Dang SM, Gerecht-Nir S, Chen J, Itskovitz-Eldor J, Zandstra PW. Controlled, scalable embryonic stem cell differentiation culture. Stem Cells . 2004 Jan;22(3):275–82.
28. Guolla L, Bertrand M, Haase K, Pelling AE. Force transduction and strain dynamics in actin stress fibres in response to nanonewton forces. J Cell Sci . 2012 Feb 1;125(Pt 3):603–13.
29. McCloy RA, Rogers S, Caldon CE, Lorca T, Castro A, Burgess A. Partial inhibition of Cdk1 in G2 phase overrides the SAC and decouples mitotic events. Cell Cycle. 2014;13(9):1400–12.
30. Schneider C a, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods . 2012;9(7):671–5.
31. Lewiner T, Lopes H, Vieira AW, Tavares G. Efficient Implementation of Marching Cubes' Cases with Topological Guarantees. J Graph Tools. 2003;8(2):1–15.
32. Slater GW, Holm C, Chubynsky M V, de Haan HW, Dubé A, Grass K, et al. Modeling the separation of macromolecules: a review of current computer simulation methods. Electrophoresis. 2009;30(5):792–818.
33. Limbach HJ, Arnold A, Mann BA, Holm C. ESPResSo—an extensible simulation package for research on soft matter systems. Comput Phys Commun. 2006;174(9):704–27.
34. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.
35. Jaalouk DE, Lammerding J. Mechanotransduction gone awry. Nat Rev Mol Cell Biol . Nature Publishing Group; 2009 Jan;10(1):63–73.
36. Gardel ML, Schneider IC, Aratyn-Schaus Y, Waterman CM. Mechanical integration of actin and adhesion dynamics in cell migration. Annu Rev Cell Dev Biol . 2010 Jan;26:315–33.
37. Gupton SL, Eisenmann K, Alberts AS, Waterman-Storer CM. mDia2 regulates actin and focal adhesion dynamics and organization in the lamella for efficient epithelial cell migration. J Cell Sci . 2007 Oct 1;120(Pt 19):3475–87.
38. Yang Y, Kulangara K, Lam RTS, Dharmawan R, Leong KW. Effects of topographical and mechanical property alterations induced by oxygen plasma modification on stem cell behavior. ACS Nano . 2012 Oct 23;6(10):8591–8.
39. Guilak F, Cohen DM, Estes BT, Gimble JM, Liedtke W, Chen CS. Control of stem cell fate by physical interactions with the extracellular matrix. Cell Stem Cell . 2009 Jul 2;5(1):17–26.
40. Kim D-H, Provenzano PP, Smith CL, Levchenko A. Matrix nanotopography as a regulator of cell function. J Cell Biol . 2012 Apr 30;197(3):351–60.

41. Oda H, Takeichi M. Evolution: structural and functional diversity of cadherin at the adherens junction. J Cell Biol . 2011 Jun 27;193(7):1137–46.

42. Li L, Bennett SAL, Wang L. 2012 Landes Bioscience . . 2012;(February):59–70.

43. Borghi N, Sorokina M, Shcherbakova OG, Weis WI, Pruitt BL, Nelson WJ, et al. E-cadherin is under constitutive actomyosin-generated tension that is increased at cell-cell contacts upon externally applied stretch. Proc Natl Acad Sci U S A . 2012 Jul 31;109(31):12568–73.

44. D'Amour K a, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol . 2005 Dec;23(12):1534–41.

45. Eastham AM, Spencer H, Soncin F, Ritson S, Merry CLR, Stern PL, et al. Epithelial-mesenchymal transition events during human embryonic stem cell differentiation. Cancer Res . 2007 Dec 1;67(23):11254–62.

46. Li D, Zhou J, Wang L, Shin ME, Su P, Lei X, et al. Integrated biochemical and mechanical signals regulate multifaceted human embryonic stem cell functions. J Cell Biol. 2010;191:631–44.

47. Shewan AM, Maddugoda M, Kraemer A, Stehbens SJ, Verma S, Kovacs EM, et al. Myosin 2 is a key Rho kinase target necessary for the local concentration of E-cadherin at cell-cell contacts. Mol Biol Cell . 2005 Oct;16(10):4531–42.

48. Weber GF, Bjerke MA, DeSimone DW. Integrins and cadherins join forces to form adhesive networks. J Cell Sci . 2011 Apr 15;124(Pt 8):1183–93.

49. Stricker J, Beckham Y, Davidson MW, Gardel ML. Myosin II-mediated focal adhesion maturation is tension insensitive. PLoS One . Public Library of Science; 2013 Jan 29;8(7):e70652.

50. Li F, Higgs HN. The mouse formin mDia1 is a potent actin nucleation factor regulated by autoinhibition. Curr Biol. 2003;13(15):1335–40.

51. Maekawa M, Ishizaki T, Boku S, Watanabe N, Fujita A, Iwamatsu A, et al. Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase. Science. 1999;285(5429):895–8.

52. Yamana N, Arakawa Y, Nishino T, Kurokawa K, Tanji M, Itoh RE, et al. The Rho-mDia1 pathway regulates cell polarity and focal adhesion turnover in migrating cells through mobilizing Apc and c-Src. Mol Cell Biol. 2006;26(18):6844–58.

53. Chaudhary A, Brugge JS, Cooper JA. Direct phosphorylation of focal adhesion kinase by c-Src: Evidence using a modified nucleotide pocket kinase and ATP analog. Biochem Biophys Res Commun. 2002;294(2):293–300.

54. Beloussov L V, Luchinskaya NN, Ermakov AS, Glagoleva NS. Gastrulation in amphibian embryos, regarded as a succession of biomechanical feedback events. Int J Dev Biol . 2006 Jan;50(2-3):113–22.

55. Beloussov L V. Mechano-geometric generative rules of morphogenesis. Biol Bull . 2012;39(2):119–26.

56. Beloussov L V. Integrating self-organization theory into an advanced course on

morphogenesis at Moscow State University. International Journal of Developmental Biology. 2003. p. 177–81.

57. Discher DE, Mooney DJ, Zandstra PW. Growth factors, matrices, and forces combine and control stem cells. Science . 2009 Jun 26;324(5935):1673–7.

58. Wang JH-C, Thampatty BP. Mechanobiology of adult and stem cells. Int Rev Cell Mol Biol. 2008;271:301–46.

59. Dard N, Louvet-Vallée S, Santa-Maria A, Maro B. Phosphorylation of ezrin on threonine T567 plays a crucial role during compaction in the mouse early embryo. Dev Biol. 2004;271:87–97.

60. Eckert JJ, McCallum A, Mears A, Rumsby MG, Cameron IT, Fleming TP. Specific PKC isoforms regulate blastocoel formation during mouse preimplantation development preimplantation development. Peptides. 2004;274:384–401.

61. Fleming TP, McConnell J, Johnson MH, Stevenson BR. Development of tight junctions de novo in the mouse early embryo: control of assembly of the tight junction-specific protein, ZO-1. J Cell Biol . 1989 Apr;108(4):1407–18.

62. Guilak F, Cohen DM, Estes BT, Gimble JM, Liedtke W, Chen CS. Control of stem cell fate by physical interactions with the extracellular matrix. Cell Stem Cell. 2009;5:17–26.

63. Khang I, Sonn S, Park J-H, Rhee K, Park D, Kim K. Expression of epithin in mouse preimplantation development: its functional role in compaction. Dev Biol. 2005;281:134–44.

64. Plusa B, Frankenberg S, Chalmers A, Hadjantonakis A-K, Moore CA, Papalopulu N, et al. Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. J Cell Sci. 2005;118:505–15.

65. Wong JY, Leach JB, Brown XQ. Balance of chemistry, topography, and mechanics at the cell–biomaterial interface: Issues and challenges for assessing the role of substrate mechanics on cell response. Surf Sci . 2004 Oct;570(1-2):119–33.

66. Zatti S, Zoso A, Serena E, Luni C, Cimetta E, Elvassore N. Micropatterning topology on soft substrates affects myoblast proliferation and differentiation. Langmuir. 2012;28:2718–26.

67. Chen CS, Mrksich M, Huang S, Whitesides GM, Ingber DE. Geometric control of cell life and death. Science (80-). 1997;276:1425–8.

68. Charest JL, Eliason MT, García AJ, King WP. Combined microscale mechanical topography and chemical patterns on polymer cell culture substrates. Biomaterials . 2006 Apr;27(11):2487–94.

69. Bettinger CJ, Langer R, Borenstein JT. Engineering substrate topography at the micro- and nanoscale to control cell function. Angew Chemie Int Ed. 2009;48:5406–15.

70. Holle AW, Engler AJ. More than a feeling: discovering, understanding, and influencing mechanosensing pathways. Curr Opin Biotechnol . Elsevier Ltd; 2011 Oct;22(5):648–54.

71. Chowdhury F, Na S, Li D, Poh Y-C, Tanaka TS, Wang F, et al. Material properties of the cell dictate stress-induced spreading and differentiation in embryonic stem cells. Nat Mater. 2010;9(1):82–8.

72. Arulmoli J, Pathak MM, McDonnell LP, Nourse JL, Tombola F, Earthman JC, et al. Static stretch affects neural stem cell differentiation in an extracellular matrix-dependent manner. Sci Rep . 2015;5:8499.

73. Irimia D, Toner M. Spontaneous migration of cancer cells under conditions of mechanical confinement. Integr Biol (Camb) . 2009 Sep;1(8-9):506–12.

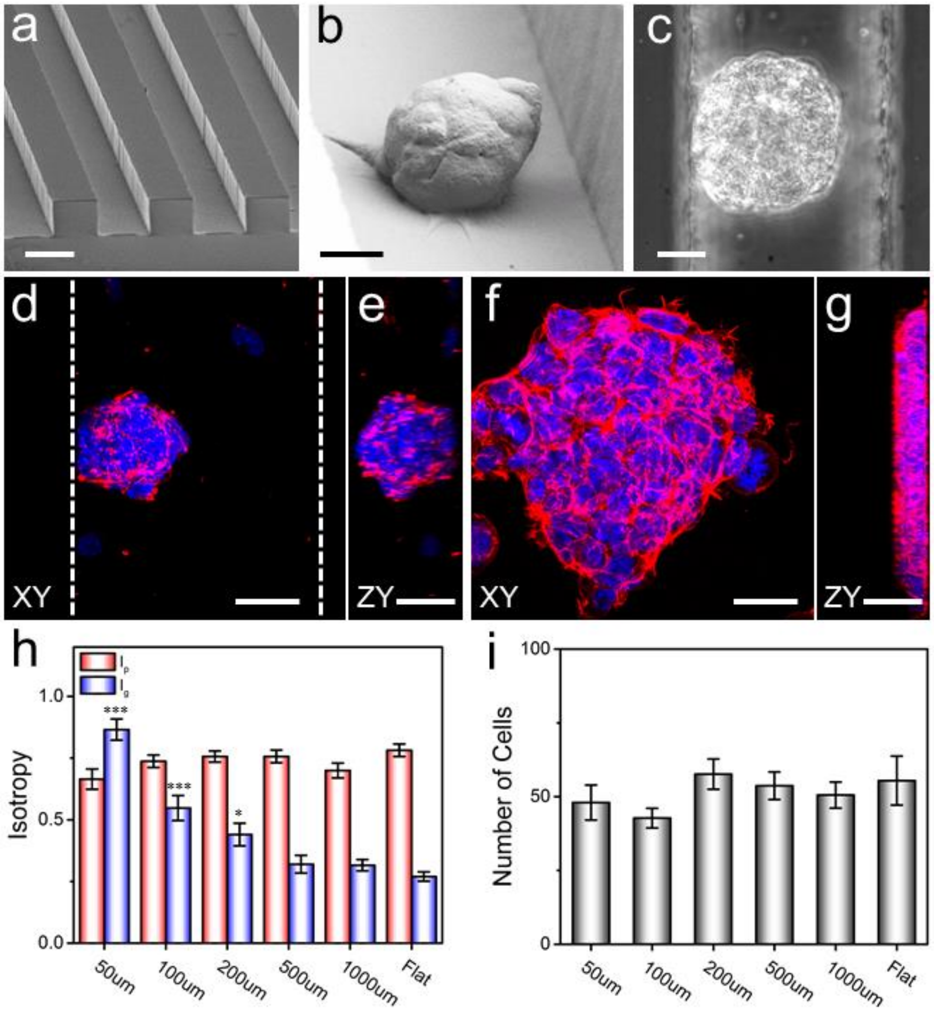


FIG. 1 Effects of topographical confinement on embryonic stem cell growth. (a) A perspective view of an SEM image of the PDMS substrate reveals the structure of a typical microtopography with 100 µm grooves and ridges. Embryonic stem cells are seeded on this 100µm groove topography for 48 hrs, whereby spheroidal aggregates begin to form as a result of the confinement properties of the microenvironment (scale bar = 100 µm). SEM (b) and phase contrast (c) images display the spheroidal geometry of the aggregates in the 100 µm groove after 48 hrs of growth (scale bars = 25 µm). Actin (red) and dapi (blue) in a top-down (d) and side (e) view further reveal the three dimensional geometry of the aggregate in a grooved space. Doted lines indicated the presence of the channel walls. Cells plated on a flat PDMS substrate demonstrated significantly different geometric shapes with their morphology resembling a circular (f, top-down) yet flat shape(g, side view) (all scale bars = 25 µm). (h) The globular isotropy Ig (blue) and planar isotropy Ip (red) which define the sphericity and circularity of the aggregate respectively revealed that at higher levels of confinement (50 µm, 100 µm, 200 µm), Ig is statistically different (*** $p<0.001$, * $p<0.05$, one way ANOVA, mean± s.e.m) to the flat substrate while Ip remains unaffected. The number of cells per aggregate (i) was also counted to ensure that the differences in morphology observed wasn't the result of cell density. The results show no statistically significant difference across any of the channels and the flat substrate. (n=25)

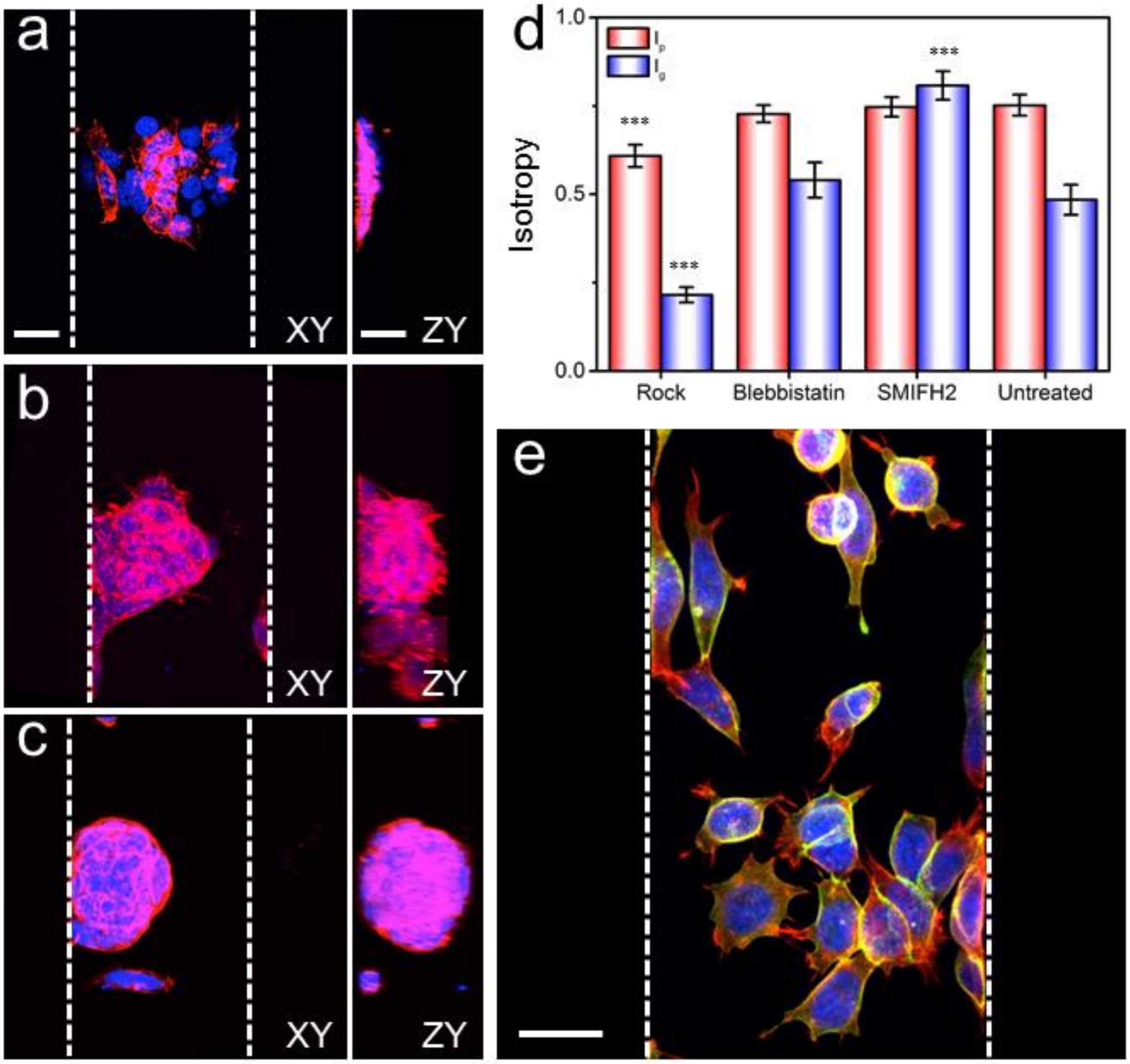


FIG. 2 The effects of inhibiting actoymyosin dynamics and E-cadherin function on spheroid formation. Max projection images of actin (red) and dapi (blue) taken by confocal microscopy displaying the effects of the drugs Y-27632 (a)(n=20), blebbistatin (b)(n=24) and SMIFH2 (c)(n=28) on embryonic stem cells grown in 100um grooved channels (scale bars in (a) = 25 μm and apply to (b) and (c)). Dotted lines represent the presence of the wall. Inhibition studies were performed to elucidate the molecular mechanisms responsible for altered development in response to confinement. (d) With the exception of blebbistatin, both Y-27632 and SMIFH2 had a significant effect on the globular isotropy Ig (blue). Planar isotropy Ip (red) appears to be less affected by the selective inhibitions with only Y-27632 demonstrating a significant effect. E-cadherin, the cell surface protein responsible for cell-cell adhesions, was blocked by exposure to a primary E-cadherin antibody prior to sending. (e) Actin (red), dapi (blue) and e-cadherin (green) display the effects of blocking e-cadherin prior to seeding in the grooves, resulting in a complete breakdown of aggregation as cells aberrantly grew in a single cell manner (scale bar = 25 μm)(n=3).

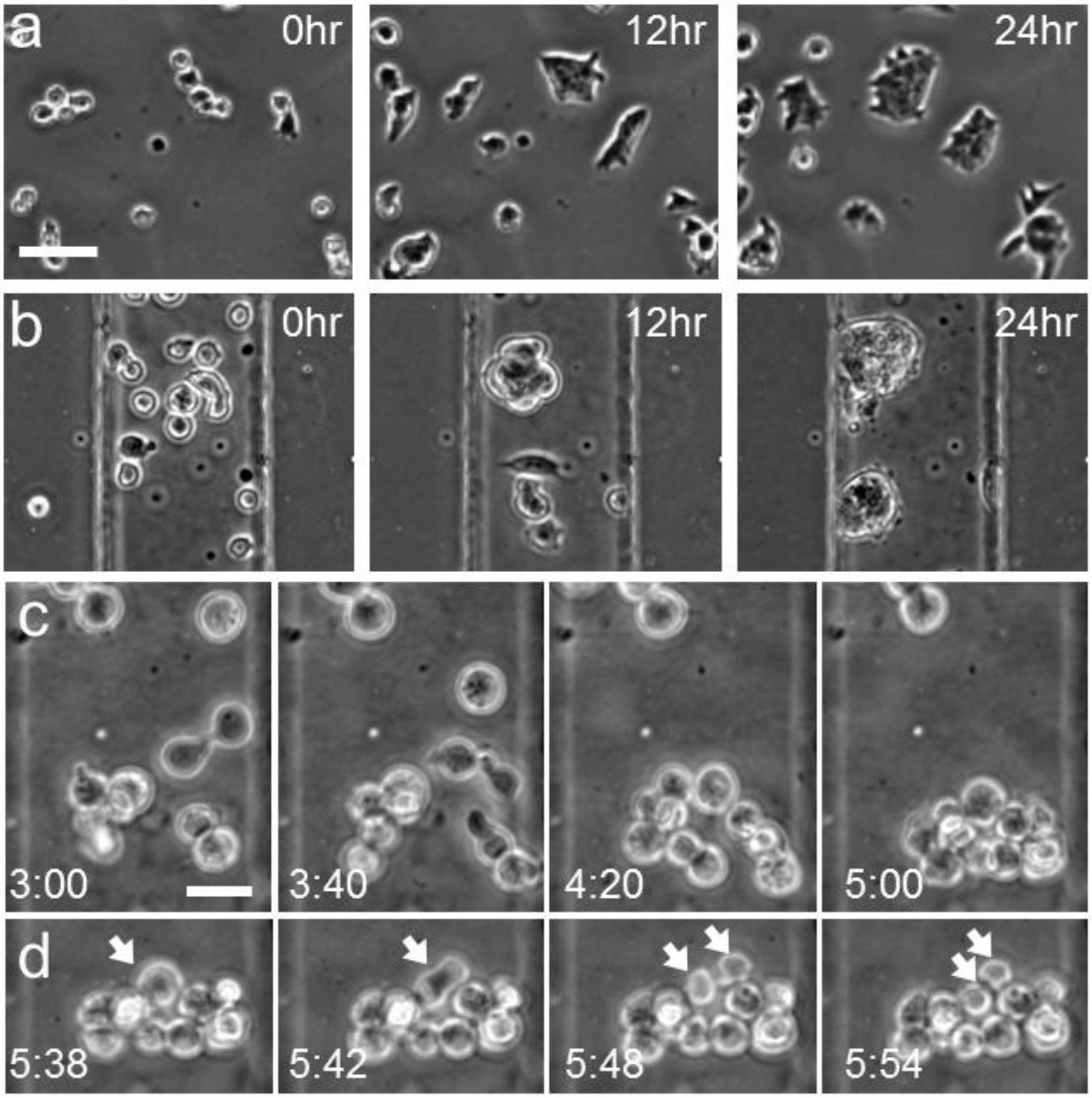


FIG. 3 Time lapse imaging of mESC proliferation and organization. Embryonic stem cells were seeded on to a flat (a) and 100μm channeled (b) PDMS topography and imaged for a 24hr period. Scale bar = 50 μm. Aggregation between cells display distinctly different formation patterns as confined cells grow in a far more three dimensional manner. (c) A closer look at the preliminary cell-cell interactions of a developing aggregate in a 100μm channel. Scattered cells diffuse until they stochastically collide eventually forming a chain of cells. This is followed by an infolding, creating a preliminary spheroidal geometry. Once aggregated, (d) growth occurs through cycles of mitosis (white arrow) and cell re-organization. Scale bar = 25 μm.

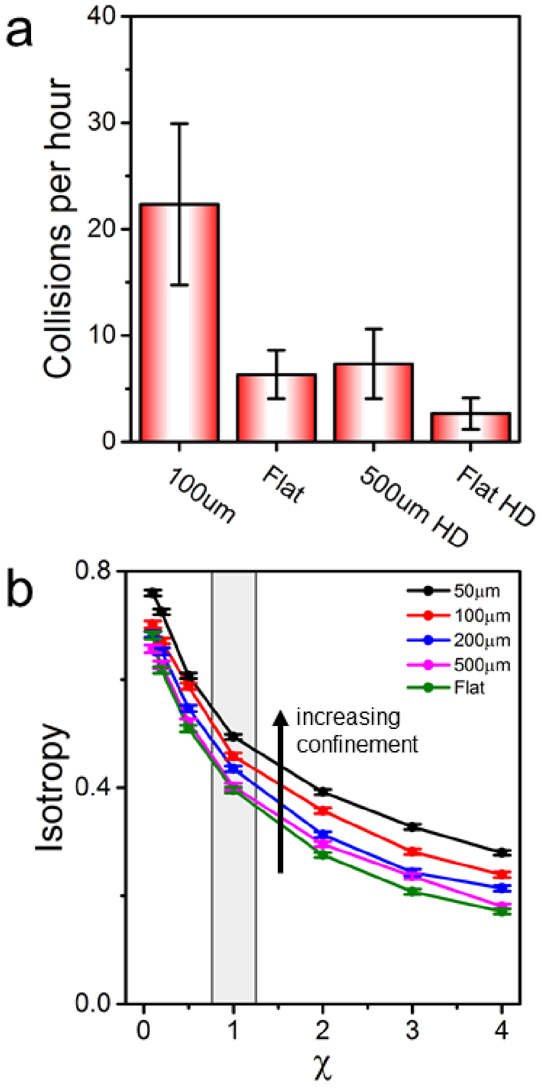


FIG. 4 Cell-cell collision dynamics and computational modeling of cellular organization and morphology. (a) The number of collision events between cells at varying levels of confinement. Cells were seeded onto PDMS topographies and recorded every 5 mins for 6hrs. To compensate for the difference in cell density due to confinement, a higher density (HD) seed matching the effective density of 100 μm channels (450 cells/mm^2) was performed for the 500 μm and flat conditions. The collision frequency of cells in 100μm channels was the only condition displaying a statistically significant difference (*** p <0.001) compared to the flat PDMS control. (b) The globular isotropy I_g of simulated spheroids at varying levels of χ (cell-substrate/cell-cell energy). Channel widths: 50 μm (black), 100 μm (red), 500 μm (blue), 1000 μm (green), flat (pink). Simulations were performed replicating experimental conditions, with a preliminary cell density of ~ 450 cells/mm^2 at varying channel widths. Cells undergo a preliminary phase of diffusion followed by cycles of duplication and relaxation. The average (n=100) I_g of aggregates demonstrates that as cell-substrate energies increase, aggregate geometry becomes flatter. At χ = 1, the simulation displays very similar results to those acquired experimentally.

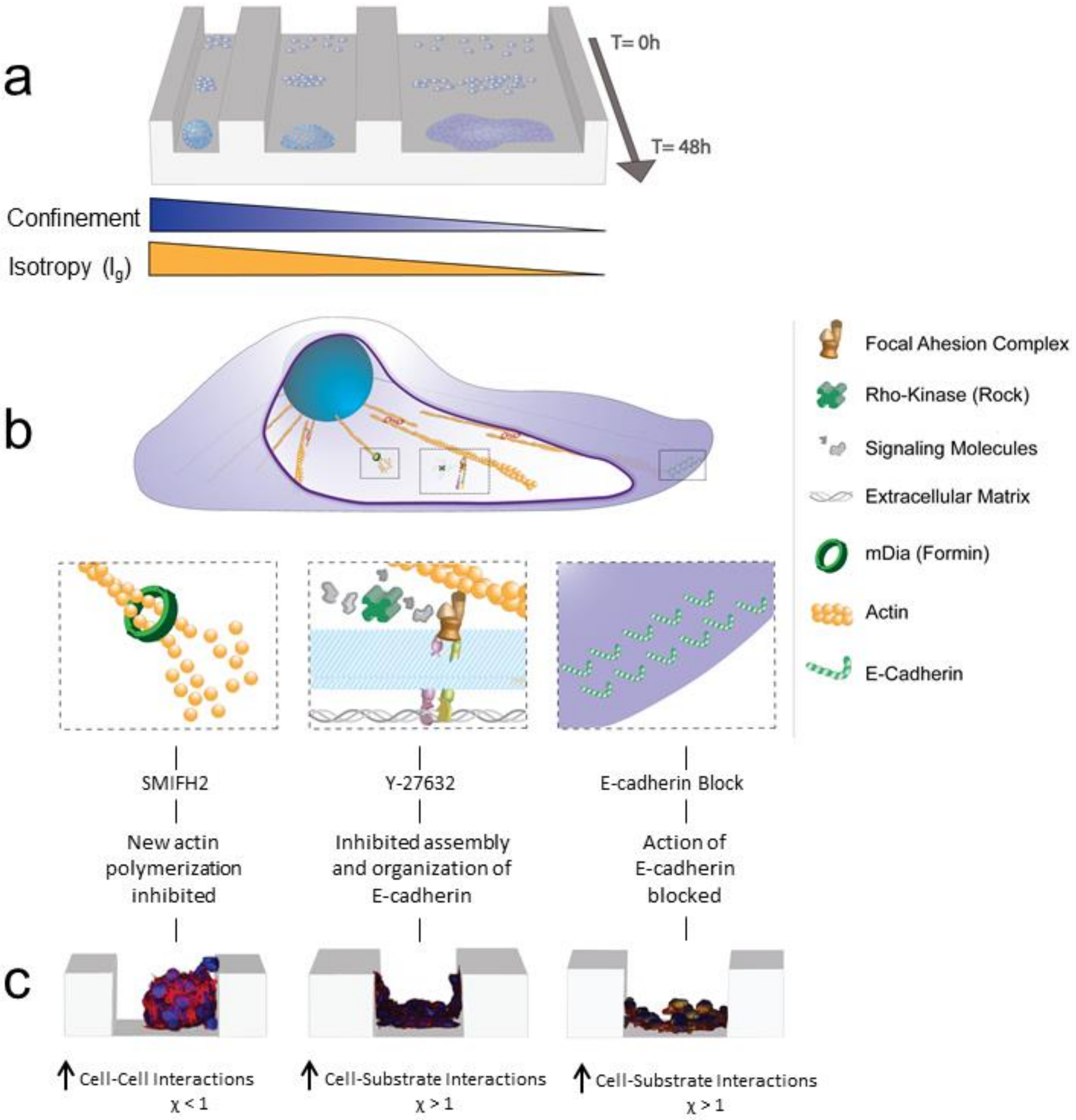


FIG. 5 Effects of confinement and selective protein inhibition in embryonic stem cell formation. (A) At T = 0hrs, cells begin to adhere to the substrate. With time, stochastic collisions induce cell-cell interactions and promote aggregate formation. Although this formation occurs in all conditions, its globular isotropy is dependent on the level of confinement imposed on the cells, with higher levels promoting spheroidal formations. (B) Cell movement and growth is dictated by multiple key proteins that regulate actin formation, contraction and cell-cell adhesion. Inhibiting mDia, and consequently actin polymerization, (C) resulted in a spheroidal formation with higher I_g values. Inhibiting Rock, and hence E-cadherin assembly, completely prevented three-dimensional development, causing cells to be highly dependent on cell-substrate adhesion. Similarly, blocking E-cadherins, the cell-cell adhering junctions, prevented cell-cell interactions and prevented any aggregation after 48 hrs.

References for Chapters 1 and 9

[1] Bustamante, C. *et al.* *Grabbing the cat by the tail: manipulating molecules one by one.* Nature Reviews Molecular Cell Biology, **1**:130–136 (2000). (Cited on page 4.)

[2] Joo, C. *et al.* *Advances in single-molecule fluorescence methods for molecular biology.* Annual Review of Biochemistry, **77**:51–76 (2008). (Cited on page 4.)

[3] De Gennes, P. *Molecular individualism.* Science, **276**:1999–2000 (1997). (Cited on page 4.)

[4] Branton, D. *et al.* *The potential and challenges of nanopore sequencing.* Nature biotechnology, **26**:1146–1153 (2008). (Cited on pages 6, 7, and 30.)

[5] Teraoka, I. *Polymer Solutions: An Introduction to Physical Properties.* Wiley (2002). ISBN 9780471389293. (Cited on pages 12, 13, 14, and 26.)

[6] Rubinstein, M. and Colby, R. *Polymer Physics.* Oxford University Press (2003). ISBN 9781613449431. (Cited on pages 15, 19, and 21.)

[7] Slater, G.W. *et al.* *Deformation, Stretching, and Relaxation of Single-Polymer Chains: Fundamentals and Examples#.* Soft Materials, **2**:155–182 (2004). (Cited on pages 17 and 28.)

[8] Marko, J.F. and Siggia, E.D. *Stretching DNA.* Macromolecules, **28**:8759–8770 (1995). (Cited on page 17.)

[9] de Haan, H.W. and Shendruk, T.N. *Force–Extension for DNA in a Nanoslit: Mapping between the 3D and 2D Limits.* ACS Macro Letters, **4**:632–635 (2015). (Cited on page 20.)

[10] Tree, D.R. *et al.* *Is DNA a Good Model Polymer?* Macromolecules, **46**:8369–8382 (2013). (Cited on page 23.)

[11] Guillouzic, S. and Slater, G.W. *Polymer translocation in the presence of excluded volume and explicit hydrodynamic interactions.* Physics Letters A, **359**(4):261–264 (2006). (Cited on pages 27 and 31.)

[12] Slater, G. *et al.* *Modeling the separation of macromolecules: A review of current computer simulation methods.* Electrophoresis, **30**:792–818 (2009). (Cited on pages 28 and 29.)

[13] Weeks, J.D. *et al.* *Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids.* The Journal of Chemical Physics, **54**:5237–5247 (1971). (Cited on page 28.)

[14] Grest, G. and Kremer, K. *Molecular dynamics simulation for polymers in the presence of a heat bath.* Physical Review A, **33**:3628–3631 (1986). (Cited on pages 28 and 29.)

[15] Sung, W. and Park, P. *Polymer translocation through a pore in a membrane.* Physical Review Letters, **77**:783 (1996). (Cited on page 30.)

[16] Muthukumar, M. *Polymer translocation through a hole.* The Journal of Chemical Physics, **111**:10371–10374 (1999). (Cited on page 30.)

[17] Chuang, J. *et al.* *Anomalous dynamics of translocation.* Physical Review E, **65**:011802 (2001). (Cited on page 31.)

[18] Gauthier, M.G. and Slater, G.W. *Nondriven polymer translocation through a nanopore: Computational evidence that the escape and relaxation processes are coupled.* Physical Review E, **79**:021802 (2009). (Cited on page 31.)

[19] Gauthier, M.G. and Slater, G.W. *A Monte Carlo algorithm to study polymer translocation through nanopores. I. Theory and numerical approach.* The Journal of Chemical Physics, **128**:065103 (2008). (Cited on page 31.)

[20] Gauthier, M.G. and Slater, G.W. *A Monte Carlo algorithm to study polymer translocation through nanopores. II. Scaling laws.* The Journal of Chemical Physics, **128**:205103 (2008). (Cited on page 31.)

[21] de Haan, H.W. and Slater, G.W. *Memory effects during the unbiased translocation of a polymer through a nanopore.* The Journal of Chemical Physics, **136**:154903 (2012). (Cited on page 31.)

[22] Sakaue, T. *Nonequilibrium dynamics of polymer translocation and straightening.* Physical Review E, **76**:021803 (2007). (Cited on page 31.)

[23] Sakaue, T. *Sucking genes into pores: Insight into driven translocation.* Physical Review E, **81**:041808 (2010). (Cited on page 31.)

[24] Saito, T. and Sakaue, T. *Dynamical diagram and scaling in polymer driven translocation.* The European Physical Journal E, **34**:1–8 (2011). (Cited on page 31.)

[25] Saito, T. and Sakaue, T. *Process time distribution of driven polymer transport.* Physical Review E, **85**:061803 (2012). (Cited on page 31.)

[26] Ikonen, T. *et al.* *Influence of non-universal effects on dynamical scaling in driven polymer translocation.* The Journal of Chemical Physics, **137**:085101 (2012). (Cited on pages 31 and 140.)

[27] Dunn, T.R. and Polson, J.M. *Evaluating the applicability of the Fokker-Planck equation in polymer translocation: A Brownian dynamics study.* The Journal of Chemical Physics, **140**:184904 (2014). (Cited on page 31.)

[28] Waugh, M. *et al.* *Interfacing solid-state nanopores with gel media to slow DNA translocations.* Electrophoresis, **36**:1759–1767 (2015). (Cited on page 138.)

[29] Shendruk, T. *et al.* *Electrophoresis: When hydrodynamics matter.* Current Opinion in Colloid & Interface Science, **17**:74–82 (2012). (Cited on page 141.)

[30] Vollmer, S.C. and de Haan, H.W. *Translocation is a nonequilibrium process at all stages: Simulating the capture and translocation of a polymer by a nanopore.* The Journal of Chemical Physics, **145**:154902 (2016). (Cited on page 141.)